Vol. II 1931

THE ANNALS

of

MATHEMATICAL STATISTICS

(Printed in U. S. A.)

Published and Lithoprinted by
EDWARDS BROTHERS, INC.

EDITORIAL COMMITTEE

A quarterly publication of the American Statistical Association, devoted to the theory and application of Mathematical Statistics

Six dollars per annum

CONTENTS OF VOLUME II

CONTENTS OF VOLUME II—Continued

Vol. II No. 1

THE ANNALS of MATHEMATICAL STATISTICS

FEBRUARY - 1931

THE RELATIONS BETWEEN STABILITY AND HOMOGENEITY*

By

L. v. Bortkiewicz

The idea of investigating the stability of statistical frequencies from the standpoint of the theory of probability goes back to the French mathematician Bienaymé From various examples taken from social and moral statistics, he was the first to establish the fact that, almost without exception, the stability in question was essentially less than the "classical norm," that is, less than the expectation which is associated with the classical scheme of independent trials with a constant underlying probability. In order to explain this discrepancy between theory and observation, Bienaymé used a modification of the traditional procedure which was characterized by the assumption that between neighboring trials in a time ordered sequence a sort of dependence existed. Though interesting in itself and among other things adopted by Cournot as his own, we shall replace this method in what follows by another, originating from Lexis, which has the advantage of a wider usefulness, in that it can be applied not only

*Translated by A R Crathorne Read before the American Statistical Association at Cleveland, Ohio. December 30, 1930.

to undulatory but to evolutory sequences.[1]

Let us assume that for a series of z successive time intervals, say years, we have found that some event (accident, death, marriage, crime) has happened x_1, x_2, ... times, and that the corresponding number of "trials," that is the numbers of persons observed, are s_1, s_2, ... so that the quotients $y_1 = \frac{x_1}{s_1}$, $y_2 = \frac{x_2}{s_2}$, ... represent a time ordered sequence of relative frequencies. Instead of assuming, as the traditional theory demands, that each term y_k of this series corresponded to a common fundamental probability p, weighted with accidental errors, Lexis assumed that each value y_k was associated with a distinct probability p_k.

As a result of this, the expected amplitude of the fluctuations of the values y_k increased, and the greater the variations in the p_k's the greater the amplitude. Under the simplifying hypothesis s_k = const. ($= s$), the corresponding standard deviation σ is defined by

$$\sigma^2 = \frac{1}{z}\sum_{k=1}^{z}(y_k - y)^2; \quad y = \frac{1}{z}\sum_{k=1}^{z} y_k$$

For the case of a constant p we may write

$$E(\sigma^2) = \frac{z-1}{z}\;\frac{p(p-1)}{s} \tag{1}$$

where E denotes "expectation." In the Lexis procedure with a variable p_k, using the notation

$$\frac{z-1}{z}\cdot\frac{p(1-p)}{s} = u^2, \quad \frac{1}{z}\sum_{k=1}^{z} p_k = p, \quad p_k - p = \varepsilon_k, \; \frac{1}{z}\sum_{k=1}^{z}\varepsilon_k^2 = \omega^2,$$

---

[1]Bienayme, in the journal "L'Institute," Vol. 7 (1831), pages 187-189, and in "Journal de la Societe de Statistique de Paris," 17e (1876), pages 199-204. A. Cournot, Exposition de la theorie des chances et des probabilities, Paris, 1843, Nos. 79 and 117

W. Lexis, "Uber die Theorie der Stabilitat statistischer Reihen," in the Jahrbuch fur Nationalokonomie und Statistik, Vol 32 (1879), pages 60 . ., reprinted in Abhandlungen zur Theorie der Bevölkesungs und Moralstatistik, Jena, 1903, pages 170-212.

the corresponding relation

$$E(\sigma^2) = u^2 + \frac{zs - z + 1}{zs}\,\omega^2 \tag{2}$$

can be derived.[1]

In the following numerical examples the numbers of observations s_k are never less than some ten thousands, while $z = 10$. Hence, as far as these and similar examples are concerned, the numerical results are not appreciably altered if, instead of (2), we use

$$E(\sigma^2) = u^2 + \omega^2 \tag{3}$$

However, a certain inaccuracy arises, if, in the application of formula (3) to the raw data, one has disregarded the fundamental assumption that s_k is constant and in the expression for u^2 has replaced s by the arithmetic mean of the z values s_k. If, however, the latter differ little from one another, such a procedure gives rise to no great discrepancy. Lexis called the quantities u and ω in formula (3) the two "fluctuation components," which combine (according to the law of composition of forces) to give the expected total fluctuation. The quantity u gives expression to the effect of the "accidental causes" in the sense of the theory of probability, and this effect grows less and less with increasing s until it vanishes for $s = \infty$. For this reason Lexis called u the normal component. He also used the term "unessential fluctuation compononent." On the other hand, ω depends on the variations of the fundamental probability, that is on the underlying general conditions, and in this sense was designated by Lexis as the physical component. We may also

[1] One does not find formula (2) in Lexis's work. He was satisfied at this point with a rather inexact method yielding an approximate result. However, this did not affect the essential part of his discussion.

call it the essential component.

The first of the two components u and ω can be easily calculated directly with sufficient approximation. The usual method is to substitute for the unknown p in the expression for u^2 the value y , the arithmetic mean of the frequencies y_k , obtaining

$$u^2 = \frac{z-1}{z} \cdot \frac{y(1-y)}{s} \tag{4}$$

As for the second component ω , it is calculated by the indirect method of substituting σ^2 for $E(\sigma^2)$ in (3) and then ω is found from $\omega^2 = \sigma^2 - u^2$. This method, however, assumes that $\sigma > u$, or what is the same thing, that the dispersion coefficient, $Q = \frac{\sigma}{u}$, is greater than 1. In his older papers, Lexis distinguished between subnormal, normal and supernormal dispersion, according to whether Q was distinctly less than 1, approximately equal to 1, or distinctly greater than 1, and found that in social and moral statistics the subnormal dispersion never occurred and the normal rarely. Supernormal dispersion was the rule. So Lexis based his scheme of a varying underlying probability on the case of supernormal dispersion In fact, from formula (3), we have

$$E(Q^2) = 1 + \left(\frac{\omega}{u}\right)^2 . \tag{5}$$

which says that the variations in the underlying probability lead us to expect values of Q greater than unity.[1]

Notwithstanding the fact that Q was usually greater than unity, Lexis did not consider this a proof that his scheme ade-

[1]Under the influence of accidental causes, Q may be less than unity not only for constant, but also for varying underlying probabilities, and this circumstance must be considered in the determination of. ω. It would carry us too far afield to go further into this matter.

quately described the actual facts. In addition to this he was more concerned with the fact that in experience Q showed a tendency to decrease with decreasing number of "trials," that is with decreasing s. Indeed, in a series of examples, Lexis had shown that a value of Q which was decidedly greater than unity when calculated for an entire country, decreased to nearly 1 when the data for the single administration districts of the same country were used. Lexis considered such behavior of Q as entirely in harmony with his scheme.

If we write formula (5) in the form

$$(6) \qquad E(Q^2) = 1 + s\,\frac{z\omega^2}{(z-1)p(1-p)},$$

we see that the excess of Q^2 over and above 1 is in expectation directly proportional to s. This was the explanation of the decrease of Q with decreasing s, for as Lexis said, we have no ground to expect that s being large or small had any bearing on the value of ω.

It is this last point about which the criticism of Lexis's dispersion theory centers Notwithstanding the endeavors of Lexis to fit his theory to statistical reality, we can show that the facts were against him as far as his assumption that ω is fundamentally independent of s is concerned If this assumption were true, then formula (6) tells us distinctly how Q decreases with diminishing s. We learn from experience that as a rule this decrease in Q is less than that given by the formula; from which it follows that the essential component, ω, has a tendency to increase with decreasing s.

If we desire to investigate just what happens in reality, a certain complication arises, because we are never able to compare groups which differ among one another as to s, but not as to p (or q). In order to eliminate to some extent the variations of p we consider the ratio of ω to p. Let $\frac{\omega}{p} = \beta$,

and call β the *relative* essential component to distinguish it from the *absolute* essential component ω. Formula (6) then becomes the following:

$$E(Q^2) = 1 + sp \frac{z\beta^2}{(z-1)(1-p)} \tag{7}$$

The product sp can be considered as the expected number of "successes." For a constant s_k ($= s$) we have

$$E(x_k) = sp_k, \qquad E\left(\frac{1}{z}\sum_{k=1}^{z} x_k\right) = sp$$

and, letting $s = \frac{1}{z}\sum s_k$, the last relation is true with sufficient approximation for a variable s_k provided the variation is not too pronounced. Let $sp = m$. Often, as in the examples which follow, p is so small that we can consider ($1-p$) as equal to 1. Formula (7) then becomes

$$E(Q^2) = 1 + \frac{z}{z-1} m\beta^2 \tag{8}$$

The question as to whether there is a connection between s and ω is now changed to an investigation of the relationship between m and β. In undertaking such an investigation empirically, we compare as to the behavior of m and β a statistical aggregate considered as a total with its component parts considered as partial aggregates. Let the number of the partial aggregates be n, and let the corresponding values of m and β as well as u, ω and σ be indicated by the subscript i, which can also serve as the ordinal number of the partial aggregate. For the total aggregate, let $i = 0$. The symbols $s_{i,k}$, $x_{i,k}$, $y_{i,k}$, $p_{i,k}$, are the s, x, y, p of the i th partial aggregate and the k th time interval. We also use

the notation

$$s_i = \frac{1}{z}\sum_{k=1}^{z} s_{i,k}\,, \quad x_i = \frac{1}{z}\sum_{k=1}^{z} x_{i,k}\,,$$

$$y_i = \frac{1}{z}\sum_{k=1}^{z} y_{i,k}\,, \quad p_i = \frac{1}{z}\sum_{k=1}^{z} p_{i,k}\,.$$

from which we have

$$s_o = \sum_{i=1}^{n} s_i\,, \qquad x_o = \sum_{i=1}^{n} x_i\,,$$

$$y_o = \frac{1}{s_o}\sum_{i=1}^{n} s_i y_i\,, \quad p_o = \frac{1}{s_o}\sum_{i=1}^{n} s_i p_i\,.$$

We have also the following relations·

$$m_i = s_i p_i\,, \quad \sigma_i^2 = \frac{1}{z}\sum_{k=1}^{z}(y_{i,k} - y_i)^2,$$

$$u_i^2 = \frac{z-1}{z}\cdot\frac{p_i(1-p_i}{s_i}\,, \quad \omega_i^2 = \frac{1}{z}\sum_{k=1}^{z} e_{i,k}^2$$

$$\text{where} \quad e_{i,k} = p_{i,k} - p_i\,, \quad \beta_i = \frac{\omega_i}{p_i}\,,$$

$$E(\sigma_i^2) = u_i^2 + \omega_i^2\,, \quad Q_i = \frac{\sigma_i}{u_i}\,,$$

and using the notation $\frac{e_{i,k}}{p_{i,k}} = \varepsilon_{i,k}$ we have further

$$\beta_i^2 = \frac{1}{z}\sum_{k=1}^{z} \varepsilon_{i,k}^2$$

Finally, corresponding to formula (8), we have

$$E(Q_i^2) = 1 + \frac{z}{z-1} m_i \beta_i^2 \tag{9}$$

We shall now apply these formulas to statistics on the frequency of suicides in Germany for the decade 1902-1911. The numbers of "trials," $s_{i,k}$, are here the populations of the regions in question; the "successes," $x_{i,k}$, are the numbers of suicides for each year. The relative frequencies, $y_{i,k}$, are found by dividing the numbers of suicides by the corresponding populations. Like various other kinds of social phenomena, the suicides in pre-war German statistics were grouped according to states, the provinces of Prussia, right Rhenish Bavaria and left Rhenish Bavaria being included as states. In this way we have forty territories of very unequal size. For the decade 1902-1911, the mean population of the territories ranged from a maximum of 6,587,000 (Rhine Province) to a minimum of 45,000 (Schaumburg-Lippe). The maximum average number of suicides per annum was 1453 (Saxony) and the minimum 7 (Schaumburg-Lippe). Corresponding to the purpose of the investigation, these suicide figures x_i, which can be considered as approximations to m_i, were arranged in descending order, with $x_1 = 1453$ and $x_{40} = 7$.

For the whole of Germany, we have $x_0 = 13173$, y_0 $214 \cdot 10^{-6}$ (that is an average number of 214 suicides per annum for each million population) The ten values $y_{0,k}$ vary between $204 \cdot 10^{-6}$ and $223 \cdot 10^{-6}$ These fluctuations are markedly greater than one expects from the classical norm. The calculation of the dispersion-quotient gives $Q_0 = 3.14$, and, as the Lexis theory demands, is greater than any one of the 40 values of Q_i.[1] These values give 2.03 as a maximum and 0.75 as a

[1]A study of suicides and of homicides in the United States yields much the same general results as those shown here for suicides in Germany. (Note by the translator.)

minimum. Fixing attention on the eight smallest values of x_i, we find an average value of 1 02 for Q_i, and of the eight values, three are larger and five less than 1. So in this example the dispersion becomes very nearly 1 by narrowing the observation field.

But we have still to find out whether Q_i decreases with x_i according to the measure of decrease that one would expect under the hypothesis that β_i is fundamentally independent of x_i. To decide this question, we let β_i = const. = β, including $\beta_0 = \beta$, and substitute also x_i for m_i in formula (9). We have then on the one hand in expected values

$$Q_0^2 = 1 + \frac{z}{z-1} x_0 \beta^2$$

and on the other hand

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{z}{z-1} \quad \frac{x_0}{n} \beta^2$$

from which follows

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{1}{n} (Q_0^2 - 1)$$

However, in our example, we find

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1.56, \qquad 1 + \frac{1}{n} (Q_0^2 - 1) = 1.22$$

and the difference 0.34 cannot be ascribed to chance for it is three times the probable error (the determination of which we cannot now take up). We must, then, assume that the average of the values β_i, for $i = 1$ to 40, is greater than β_0. Why this is so we shall see in the following discussion.

We consider now the mutual relationship between the deviations $\varepsilon_{i,k}$ and $\varepsilon_{j,k}$ which refer to two arbitrary territories N_i and N_j, and we build up according to the formula for a

correlation coefficient the expression

$$\gamma_{i,j} = \frac{1}{z} \sum_{k=1}^{z} \frac{\varepsilon_{i,k}\, \varepsilon_{j,k}}{\beta_i \beta_j}$$

The number of combinations of the subscripts i and j is $\frac{n(n-1)}{2}$, so there are that many values $\gamma_{i,j}$. Finally we construct a weighted arithmetic mean of these values according to the formula,

$$\gamma = \frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} m_i m_j \beta_i \beta_j \gamma_{i,j}}{\sum_{i=1}^{n} \sum_{j=i+1}^{n} m_i m_j \beta_i \beta_j}$$

The expression γ serves to characterize the mutual relationship of time ordered series of fundamental probabilities $p_{i,k}$, hence also of relative frequencies $y_{i,k}$, which may be considered as approximations to $p_{i,k}$. If we give the name "syndromy" to such an array of simultaneously distinct fundamental probabilities (or relative frequencies), we may call γ a "coefficient of syndromy." For $\gamma = 1$, we shall speak of "isodromy," for $1 > \gamma > 0$, of "homodromy," for $\gamma = 0$, of "paradromy," and for $\gamma < 0$, of "antidromy." We may include the last three cases, namely $\gamma < 1$, under the name "anisodromy."

With the help of γ we can exhibit the relation between β_0 on the one hand and the n values $\beta_1, \beta_2, \ldots, \beta_n$ on the other hand as follows:

$$(10) \quad m_0^2 \beta_0^2 = \sum_{i=1}^{n} m_i \beta_i^2 + \gamma \left\{ \left(\sum_{i=1}^{n} m_i \beta_i \right)^2 - \sum_{i=1}^{n} m_i^2 \beta_i^2 \right\}$$

Since $m_{\bullet}=\sum_{i=1}^{n} m_i$, we find for $\gamma = 1$, from (10)

(11) $$\beta_{\circ}=\frac{\sum_{i=1}^{n} m_i \beta_i}{\sum_{i=1}^{n} m_i}$$

and for $\gamma < 1$

(12) $$\beta_{\circ}<\frac{\sum_{i=1}^{n} m_i \beta_i}{\sum_{i=1}^{n} m_i}$$

Hence, only in the case of isodromy is the assumption justified that the relative essential fluctuation component for the total aggregate is as large as that for the partial aggregates. In every other case, namely for anisodromy, the relative essential component for the total aggregate falls below the level for the partial aggregates more and more as γ becomes less and less.

In the suicide example under consideration we have homodromy, which is reasonable, since the fluctuations in suicide frequency in the single states are influenced in part by factors which are not local but general for all Germany. Somewhat tedious calculations give $\gamma = 0.38$. At the same time we find $\beta_{\bullet} = 0.0246$ approximately, while the average for β_i, $i = 1$ to 40 is 0.0392.

If now we group the 40 states into five groups so that states numbered 1 to 8 form the first group, states numbered 9 to 16 the second, and so on, we find as average values of β_i , 0.0354, 0.0358, 0.0485, 0.0528 and 0.0767. The quantities β_i then show a tendency to increase as x_i (or m_i) decreases.

If, as in this example, the total aggregate is a "natural unit," we should expect to have homodromy in the vast majority of cases. On the other hand, we should expect paradromy if the total aggregate is an "artificial unit," that is, one made up by

throwing together entirely unrelated groups. As an illustration of paradromy we take the array of marriage frequencies for the six cities, Barcelona, Birmingham, Boston, Leipzig, Melbourne and Rome, for the decade 1899-1908. By marriage frequence we mean the ratio of the number married (twice the number of marriages) to population.

For the six cities taken as a whole, with a total population of about three million, the marriage frequence $y_{o,k}$ varies between 18.00 and 19.02 per cent with an average of 18.38 per cent. The dispersion coefficient Q_o is 3.17. For the six cities taken singly in the above order, each with a population of about half a million, the values of Q_i are 2.69, 4.32, 4.17, 2.88, 3.76 and 2.72, with an average 3 42, somewhat higher than Q_o This result is a direct contradiction of the statement of Lexis that a narrowing field of observation reduces the value of Q. Lexis, without giving the matter much thought, worked with the hypothesis that isodromy, or at least a decided homodromy, always existed. In our example, however, we have paradromy, if not antidromy, for we find γ to be -0.054. Corresponding to this, we have β_o less than each of the values β_1 to β_6, for β_o approximates 0.0167 while β_i, $i = 1$ to 6, lies between 0 0334 and 0.0563. The quadratic mean of these quantities is 0.0450.

It is of prime interest to investigate for paradromy the theoretical relation of β_o to the quadratic mean of the values β_1, β_2 . . . β_n and of Q_o to the quadratic mean of Q_1, Q_2 . . Q_n, for the case m_i = const. = m. In this case, $m_o = nm$, and if O is substituted for γ in (10) we have

$$\beta_o^2 = \frac{1}{n^2}\sum_{i=1}^{n}\beta_i^2 \quad \text{, whence} \quad \beta_o = \frac{1}{\sqrt{n}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\beta_i^2}$$

At the same time, we find on the one hand, from (9), the ex-

pected value

$$Q_0^2 = 1 + \frac{z}{z-1} m_0 \beta_0^2 ,$$

or

$$Q_0^2 = 1 + \frac{z}{z-1} \frac{m}{n} \sum_{i=1}^{n} \beta_i^2$$

and on the other hand

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{z}{z-1} \frac{m}{n} \sum_{i=1}^{n} \beta_i^2$$

whence

$$Q_0 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} Q_i^2}$$

In the marriage frequence example, where the quantities m_i, though not equal, differ very little from one another, we have the values already found

$$\beta_0 = 0.0167 \text{ and } \qquad Q = 3.17$$

to compare with the values

$$\frac{1}{\sqrt{n}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \beta_i^2} = 0.0184$$

and

$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} Q_i^2} = 3.49$$

The differences $0.0167 - 0.0184 = -0.0017$ and $3.17 - 3.49 = -0.32$ are explained partly by the fact that the assumption $m_i = \text{const.}$ is not exactly in accord with the facts, and partly because para-

dromy is really not present as assumed, but only a weak antidromy. This last should, however, be considered as due to chance. The artificial character of a total aggregate shows itself in paradromy.

Of the two quantities Q and β, only the latter can be considered as a proper measure of the stability of a statistical frequency—more exactly, of the corresponding fundamental probability. And, since on account of formulas (11) and (12), the total aggregate can never show a higher value of β than the average for the partial aggregates (because the upper limit for γ is 1), we obtain a glimpse of the question of the connection between stability and homogeneity.

The idea of homogeneity as we here understand it has reference to the result of the decomposition of a statistical aggregate according to some attribute or complex of attributes. The aggregate may consist of s elements, say s human beings and the decomposition may yield N sub-aggregates containing s', s'' . . . elements Let some event A be observed x times in the total aggregate and x', x'', . . . times in the sub-aggregates. If we find the relative frequencies

$$y = \frac{x}{s}, \qquad y' = \frac{x'}{s'}, \qquad y'' = \frac{x''}{s''}, \ldots\ldots$$

then, on account of the two identities, $s' + s'' + \ldots = s$, and $x' + x'' + \ldots = x$, we have the relation

$$y = \frac{s'y' + s''y'' + \ldots}{s' + s'' + \ldots}$$

The "general frequency" then appears as the weighted arithmetic mean of the "special frequencies," y', y'', . . .

The theory of probabilities, with more or less assurance, furnishes us a criterion for deciding whether or not the deviations of the quantities y', y'', . . . from y are due to chance.

If they are not due to chance we say that the total aggregate "reacts" to the decomposition in question and that the attribute or complex of attributes which governs the decomposition is "relevant." If they are due to chance, we say that the total aggregate does not react to the decomposition and that the attribute is "indifferent"

According to the standpoint of the theory of probability, the relative frequencies y, y', y'' . . . as also the quotients $\frac{S'}{S}$, $\frac{S''}{S}$, . . . can be considered as approximations of distinct probabilities. If we designate the two series of probabilities thus inferred by p, p', p'', . . . and g', g'', . . . respectively, we find

$$(13) \qquad p = g'p' + g''p'' + \cdots$$

and the character of the attribute in question as relevant or indifferent finds expression in the fact that the "special probabilities" p', p'', . . . either differ from one another or are all equal to p, the "general probability"

For every ample enough complex of attributes we can imagine the decomposition going on and on by applying one attribute of the complex after another Finally a point is reached where the sub-aggregates no longer react to further decomposition, or, expressed otherwise, the supply of relevant attributes is exhausted, and the probabilities p', p'', . . . which are associated with these sub-aggregates are called "elementary probabilities." In this case we say that the sub-aggregates themselves are "completely homogeneous" with reference to the event A.

The total aggregate—still in reference to A —is the more diversified the more the elementary probabilities p', p'', . . . differ among themselves, that is, the more they differ from p. It is reasonable to take as a measure of this diversity the expression

δ , defined by

$$(14) \qquad \delta^2 = g'(p'-p)^2 + g''(p''-p)^2 + \cdots$$

Diversity and homogeneity are antithetical notions; the more undiversified the aggregate, the more it is homogeneous, and vice versa.

In order to apply this view of homogeneity, now considered for itself, to the procedure and the examples which we have brought forward in the discussion of stability, we must disregard the time fluctuations of the probabilities in question. That is, we do not use the quantities $p_{i,k}$ but fix attention on the probabilities p_i which refer to an individual time interval of n partial intervals—say a decade. By carrying out repeatedly the decomposition according to formula (13), the quantities p_i , p_0 not included may be expressed in the form

$$p_i = g_i' p_i' + g_i'' p_i'' + \cdot$$

where p_i' , p_i'' . . are elementary probabilities. Corresponding to formula (14), we have

$$(15) \qquad \delta_i^2 = g_i'(p_i'-p_i)^2 + g_i''(p_i''-p_i)^2 + \cdot$$

If we designate the proportion of the i th partial aggregate to the total aggregate by c_i , that is, if we let $\frac{s_i}{s_0} = c_i$, we find

$$p_0 = \sum_{i=1}^{n} c_i p_i$$

and at the same time

$$(16) \qquad \delta_o^2 = \sum_{i=1}^{n} \left\{ c_i g_i' (p_i' - p_o)^2 + c_i g_i'' (p_i'' - p_o)^2 + \cdot \right\}$$

The number of summands in (16) is nN, since there are n partial aggregates and each of these is a totality of N sub-aggregates. It may easily occur that some of the nN elementary probabilities are equal and this is expected in connection with elementary probabilities which are associated with similar sub-aggregates. But even in the most extreme case, where the elementary probabilities are equal without exception, we cannot say that the probabilities p_i are all alike. This can occur only when the values g_i', g_i'', ... are independent of i. This highly improbable case is excluded from our discussion. We have then

$$(17) \qquad \sum_{i=1}^{n} c_i (p_i - p_o)^2 > 0$$

From (15) and (16), we have the following:

$$g_i' (p_i' - p_o)^2 + g_i'' (p_i'' - p_o)^2 + \cdot \quad = \delta_i^2 + (p_i - p_o)^2$$

$$\delta_o^2 = \sum_{i=1}^{n} c_i \delta_i^2 + \sum_{i=1}^{n} c_i (p_i - p_o)^2$$

so that, on account of (17)

$$\delta_0^2 > \sum_{i=1}^{n} c_i \, \delta_i^2$$

and *a fortiori*

$$(18) \qquad \delta_0 > \sum_{i=1}^{n} c_i \, \delta_i$$

The total aggregate is then under all circumstances less homogeneous than the partial aggregates are on the average

This statement might possibly correspond to the every-day meaning of the word "homogeneity," which carries with it no precise quantitative idea. Indeed, when we consider that in the case of the total aggregate we have to take into account not only the lack of homogeneity within the partial aggregates, but also the diversity with which the partial aggregates may make up the whole, we are inclined to say that the total aggregate is less homogeneous than any of its parts. With that idea, however, we do not hit upon the right thing as far as our mathematical criterion of homogeneity is concerned. The inequality (18) says only, that the average of the values δ_1, δ_2, . . . δ_n is less than δ_0, not that each one is less than δ_0.

In our foregoing discussion of stability as measured by the relative essential fluctuation component, we found that for the total aggregate the stability was higher than the average for the partial aggregates, except for the case of isodromy, which in practice rarely occurs. Hence, there exists between homogeneity and stability an antagonistic relation—small homogeneity goes hand in hand with great stability. For example, the provinces into which a country may be divided will show, on the average, a greater homogeneity and at the same time a lesser stability in reference to an event A than will the country taken as a whole.

Again, the districts into which the provinces may be divided will on the average show a greater homogeneity associated with a still smaller stability. We can say that in general the homogeneity increases with the narrowing of the field of observation, while the stability decreases.

Is this to be considered as a warning against the all too popular diversification of statistical material which is being more and more accepted in research methods? Not in the least. That would be an obsolete point of view, as if the problem of statistics consisted in a search for most stable values. Rather does the opposition between homogeneity and stability give direction to business practice, especially to that branch of business which is in such close touch with statistics, namely insurance, where stability is of prime importance. It has been known for a long time that it contributes to the even tenor of the business side if the risks are as heterogeneous as possible. It is of advantage if the insured persons or things are spread relatively widely according to geographical and other points of view, instead of concentrating on a limited territory or few kinds of risks.

Accordingly, even if this thesis, that an antagonistic relation exists between homogeneity and stability, seems surprising and strange, we find on closer consideration that the theory agrees with a practice which has instinctively grasped the true situation. It is now twelve years since I had the first opportunity to explain at greater length than here the foregoing developed ideas and with the verifying data to present them to my colleagues.[1] As far as I know, only one of these has taken a definite stand in the matter. This is John Maynard Keynes[2] He makes the charge against me, that instead of clearing up a very simple matter, I have befogged it with a profusion of mathematical formulas

[1]Homogeneitat und Stabilitat in der Statistik, in the Skandinavisk Aktuarietidskrift, 1918, pages 1-81, Upsala.

[2]A treatise on probability, London, 1921, pages 403-405.

and new technical terms, and he believed that he could show this best by an example of my own from the field of insurance. In referring to this example, Keynes thought that the distinction made by myself in a much earlier publication between a general probability p and the special probabilities $p_1, p_2, \ldots$ was the one in question, where

$$p = \frac{z_1}{z} p_1 + \frac{z_2}{z} p_2 + \cdots$$

Keynes further expressed himself as follows:

> "If we are basing our calculations on p and do not know p_1, p_2, etc., then these calculations are more likely to be borne out by the result if the instances are selected by a method which spreads them over all the groups 1, 2, etc., than if they are selected by a method which concentrates them on group 1. In other words the actuary does not like an undue proportion of his cases to be drawn from a group which may be subject to a common relevant influence *for which he has not allowed.* If the *a priori* calculations are based on the average over a field which is not homogeneous in all its parts, greater stability of result will be obtained if the instances are drawn from all parts of the non-homogeneous total field. than if they are drawn now from one homogeneous subfield and now from another This is not at all paradoxical. Yet I believe, though with hesitation, that this is all that Von Bortkiewicz's elaborately supported mathematical conclusion amounts to."

Suppose, for example, that a fire insurance company insures

[1]Here z refers to a series of "equally likely events," which is broken up into groups of $z_1, z_2, \ldots$ equally likely events. Hence $z = z_1 + z_2 + \ldots$.

two kinds of buildings, dwellings and factories, which are classified as different grades of fire risks, for insurance premiums which are not graded. The premium is to be calculated per unit on the supposition that the risks in the two categories are divided in a definite proportion Then, according to Keynes, a greater stability in the business is guaranteed if every year dwellings as well as factories are insured, than if in one year only dwellings and in another year only factories are insured. This is certainly true and requires no lengthy argument. But it has nothing whatever to do with my thesis of the antagonistic relation between stability and homogeneity.

To give an example which does illustrate my theory, think of three insurance companies, A, B, and C A insures only dwelling houses, B only factories, while C insures both. The premiums in A, B, and C are different because of the different classes of risks. It is assumed in C that there is no grading of premiums A premium per unit is charged which is calculated according to the relative number of the two risks. The premium is to be just high enough so that for a period of years, allowing for variations due to chance, the damages are just covered. In the course of this period, the danger of fire varies from year to year, showing gains in some years, losses in others. Such fluctuations of fire hazard would correspond in my scheme to the variations of the probabilities $p_{i,k}$ with respect to k, while $p_{i,k}$ is associated with A, $p_{2,k}$ with B, and $p_{0,k}$ with C. And in accord with my theory that, except in the case of isodromy, the values $p_{0,k}$, relatively speaking, show weaker variations than $p_{1,k}$ and $p_{2,k}$ do on the average, the insurance company C would show relatively smaller fluctuations of fire damage from one year to another, resulting in a more stable business than would be shown by the average of A and B. The mixed character of the risks would be conducive to greater stability. In the case of C a certain compensation of effects would take place

which the time variations of the two-sided fundamental probabilities would make manifest on the business side.[1] But Keynes says nothing of these variations. He simply missed the point of my argument and his remarks were not relevant.

It is to be hoped that the new exposition of my theory, although, or because, it is essentially shorter than the older one, will give no cause for a similar misunderstanding.

[1]This compensation would also appear in the more complicated case where the proportions of the risks in C are not unchangeable as is assumed in the text, but would change from year to year (the premium being adjusted accordingly). We need not go further into this matter because, in my theory, the composition of $s_{\cdot,k}$ out of the component parts $s_{i,k}$ is considered as fixed. In my examples, this composition varied, but the fluctuations were insignificant in comparison to the variations of the values $p_{i,k}$ See Skandinavisk Aktúarietidscrift, pages 69-70.

L. v. Bortkiewicz.

BAYES' THEOREM

An Expository Presentation*

By

EDWARD C MOLINA

American Telephone and Telegraph Company

Bayes' theorem made its appearance as the ninth proposition in an essay which occupies pages 370 to 418 of the Philosophical Transactions, Vol. 53, for 1763 An introductory letter written by Richard Price, "Theologian, Statistician, Actuary and Political Writer,"[1] begins thus:

> "I now send you an essay which I have found amongst the papers of our deceased friend, Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved."

A few lines further on Price says:

> "In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know

*Read before the American Statistical Association during the meeting of the American Association for the Advancement of Science in Cleveland, Ohio, December, 1930.

[1] These titles are associated with the name of Price in the frontispiece portrait of him bound with the December, 1928, issue of *Biometrika*.

nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times."

. . . .

"Every judicious person will be sensible now that the problem mentioned is by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to assure a foundation for all our reasonings concerning past facts, and what is likely to be hereafter."

No one will dispute the importance ascribed to Bayes' problem by Price, in fact, a paper by Karl Pearson on an extension of Bayes' problem is entitled "The Fundamental Problem of Practical Statistics" Opinions differ, however, as to the validity and significance of the solution submitted in the essay for the problem in question. In view of this situation I shall limit myself today to an exposition of the fundamental characteristics of the problem Bayes' theorem deals with and shall give no consideration to its interesting applications.

The exposition may be outlined as follows: after specifying the class of problems to which Bayes' theorem pertains, I shall:

I. Discuss briefly two problems, each of which will emphasize one of two kinds of *a priori* probabilities which should be constantly borne in mind when Bayes' theorem is under consideration,

II. Partially analyze a certain ball-drawing problem which will not only serve as an introduction to the algebra of Bayes' theorem but will later help to throw light on its significance,

III. Present Bayes' problem and the related theorem.

IV. Make some remarks on the value of the theorem and the controversies which it raised.

In carrying out this plan I shall find it convenient to ignore the historic order of events.

When probability is the subject under consideration one an-

ticipates problems such as: A coin is about to be tossed 15 times; What is the probability that heads will turn up seven times? A sample of 100 screwdrivers is to be taken from a case containing 1000 screwdrivers of which 300 are known to be defective. what is the probability that the sample will contain 25 defectives?

These are direct, or *a priori, probability problems.* In each of them the nature of a game, or an experiment, is specified in advance and then a question is asked relating to one, or more, of the possible outcomes of the game or experiment Problems of this type have occupied the attention of mathematicians since the days of Pascal and Fermat, the creators of the mathematical theory of probability.

An inverse class of problems of great practical significance, called *a posteriori* probability problems, came into prominence with the publication of Bayes' essay. In these we find specified the result or outcome of a game which has been played, whereas the question then asked is whether the game actually played was one or some other of several possible games This type of problem is usually stated as follows:

> "An event has happened which must have arisen from some one of a given number of causes; required the probability of the existence of each of the causes"

I

Consider this example: During his sophomore year Tom Smith played on both the baseball and football varsity teams; we have been informed that he broke his ankle in one of the games; what are the *a posteriori* probabilities in favor of baseball and football, respectively, as the baneful cause of the accident?

Evidently the answer depends on the number of baseball and football games played during their respective seasons and also on the likelihood of a man breaking an ankle in one or the other of

these two games. As a concrete case assume that:

1. At Smith's college an equal number of baseball and football games are played per season;
2. Statistical records indicate that if a student participates in a baseball game the probability is 2/100 that he will break an ankle and that, likewise, the probability is 7/100 for the same contingency in a football game.

In view of the first of these two assumptions our conclusions as to the cause of the accident may be based entirely on the information contained in the second assumption. The odds are two to seven, so that the *a posteriori* probabilities regarding the two admissible causes are.

For baseball, $2/(2+7) = 2/9$
For football, $7/(2+7) = 7/9$.

Now consider this other example. A lone diner amused himself between courses by spinning a coin. We elicited from the waiter that in 15 spins heads turned up seven times Moreover, from our point of observation, the size of the coin indicated that it was either a silver quarter or a ten-dollar gold piece. What are the *a posteriori* probabilities in favor of the silver quarter and the gold piece, respectively?

If the lone diner were a professor from one of our eastern universities we would not hesitate a moment in declaring that the coin spun was a quarter But it happens that the gentleman was a member of the Cleveland Chamber of Commerce. dining at the Bankers' Club. We must, therefore, give the matter more careful consideration. The number of quarters and gold pieces usually carried by a banker and the probabilities of obtaining the observed result by spinning coins are relevant; let us assume, therefore that:

1. The small change purse of a Cleveland financier contains, on the average, ten-dollar gold pieces and quarters in the ratio of

eight to three.

Moreover, we may assume (in fact we know) that:

2. If either a quarter or a gold piece is spun 15 times, the probability that heads will turn up seven times is approximately 1/5.

The second of these two items of information makes the *a posteriori* probabilities depend entirely on the first item. Clearly the odds are eight to three and we conclude:

For a quarter, *a posteriori* probability=3/(3+8) = 3/11.
For a goldpiece, *a posteriori* probability=8/(3+8) = 8/11.

Now regarding the general *a posteriori* problem,

> "An event has happened which must have arisen from some one of a number of causes; required the probability of the existence of each of the causes,"

what do the two examples we have just considered suggest? In both problems we inquired into:

1. The frequency with which each of the possible causes is met BEFORE THE OBSERVED EVENT HAPPENED. This frequency is called the *a priori existence* probability for the corresponding cause.
2. The probability that a cause, if brought into play, would reproduce the observed event. This probability will hereafter be referred to as the *a priori productive* probability for the cause in question.

In the case of the broken ankle, the *a priori existence* probabilities were equal and took no part in our conclusion; we based the *a posteriori* probabilities entirely on the *a priori productive* probabilities. We did just the opposite with reference to the coin spun by the Cleveland financier, on account of the equality of the *a priori productive* probabilities we deduced *a posteriori* prob-

abilities in terms of the unequal *a priori* existence probabilities.

It is apparent that our two examples represent extreme cases. In general, the solution of an inverse or *a posteriori* problem, involving a number of causes, one of which must have brought about a certain observed event, depends on both sets of direct, or *a priori* probabilities. Those of the first set give the frequency with which the various causes were to be expected before the observed result occurred, those of the second set give the frequencies with which the observed result would follow from the various causes if each were brought into play.

II

Bearing in mind the two distinctly different sets of *a priori* probabilities required in arriving at *a posteriori* conclusions regarding the possible causes of an observed event, we must now give some thought to the algebra of the subject before taking up Bayes' problem and theorem. For this purpose consider the following bag problem:

A bag contained M balls, of which an unknown number were white. From this bag N balls were drawn and of these T turned out to be white. What light does this outcome of the drawings throw on the unknown ratio of the number of white balls to the total number of balls, M, in the bag? Let x be this unknown ratio.

Two cases of this problem may be considered:

Case 1.—After a ball was drawn it was replaced and the bag was shaken thoroughly before the next drawing was made.
Case 2.—A drawn ball was not replaced before the next drawing.

These two cases become essentially identical when the total number of balls in the bag is very large compared with the number drawn. Case 1 will serve as an introduction to Bayes' prob-

lem; later we will find it highly desirable to consider Case 2.

We are confronted with $(M+1)$ possible hypotheses or causes before the drawings took place:

1 – the unknown value of x is $x_0 = 0/M$,
2 – the unknown value of x is $x_1 = 1/M$,
3 – the unknown value of x is $x_2 = 2/M$,

. . . . . .

$k+1$ – the unknown value of x is $x_k = k/M$,

. . . . . .

$M+1$ – the unknown value of x is $x_M = M/M = 1$.

Let $w(x_k)$ be the *a priori* existence probability for the k'th hypothesis; by this is meant the probability in favor of the k'th hypothesis based on whatever information was available regarding the contents of the bag prior to the execution of the drawings.

Let $B(T, N, x_k)$ be the *a priori productive* probability for the k'th hypothesis; by this is meant the probability of obtaining the observed result (T whites in N drawings) when the value of x is k/M.

Then, the *a posteriori* probability, or probability after the observed event, in favor of the k'th hypothesis is

$$(1)^1 \qquad P_k = \frac{w(x_k)\, B(T, N, x_k)}{\sum_{k=0}^{M} w(x_k)\, B(T, N, x_k)}$$

For Case 1 of our bag problem we have

$$B(T, N, x_k) = \binom{N}{T} x_k^{\,T} (1-x_k)^{N-T}$$

---

[1] This is the Laplacian generalization of Bayes' formula, although in some textbooks it is referred to as "Bayes' Theorem." A relatively short demonstration of it is given by Poincaré in his *Calcul des Probabilités*. See also Fry, *Probability and its Engineering Uses*, Art. 49.

where $\binom{N}{T}$ represents the number of combinations of N things taken T at a time. Substituting in (1), we obtain, after canceling from numerator and denominator the common factor $\binom{N}{T}$,

$$(2)\qquad p_k = \frac{w(x_k)\, x_k^{T}\, (1-x_k)^{N-T}}{\sum_{k=0}^{M} w(x_k)\, x_k^{T}\, (1-x_k)^{N-T}}$$

If in equation (2) we give k successively the values a, $a+1$, $a+2$, . . . $b-1$, b and add the results, we have

$$p_a + p_{a+1} + \cdots + p_b$$

or

$$(3)\qquad p(x_a x_b) = \frac{\sum_{k=a}^{k=b} w(x_k)\, x_k^{T}\, (1-x_k)^{N-T}}{\sum_{k=0}^{M} w(x_k)\, x_k^{T}\, (1-x_k)^{N-T}}$$

for the *a posteriori* probability that the unknown ratio of white to total balls in the bag lies between a/M and b/M, both inclusive.

III

BAYES' PROBLEM

Consider the table represented by the rectangle $ABCD$ in Fig. 1. On this table a line OS was drawn parallel to, but at an unknown distance from, the edges AD and BC. Then a ball was rolled on the table N times in succession from the

edge AD toward the edge BC. As indicated in the figure it was noted that T times the ball stopped rolling to the right of the line OS and $N-T$ times to the left of that line.

What light does this information shed on the unknown distance from AD to OS? In more technical terms, what is the *a posteriori* probability that the unknown position of the line OS lies between any two positions in which we may be interested?

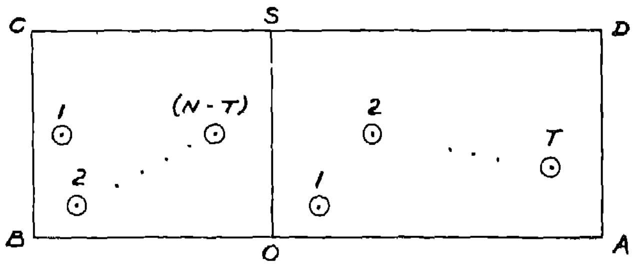


Fig. 1.

Each rolling of the ball was executed in such a manner that the probability of the ball coming to rest to the right of OS is given by the unknown ratio of the distance OA to the length BA of the table; likewise, the probability of the ball stopping to the left of OS is given by the ratio of the distance BO to the length BA.

$$\text{Set} \quad x = OA/BA, \quad 1-x = BO/BA.$$

The only difference between this problem and the bag of balls problem is that now the possible values of x are not restricted to the finite set $0/M, 1/M, 2/M, \ldots (M-1)/M, M/M$; in the table problem x may have had any value whatever between the limits of 0 and 1. Therefore equation (3) will answer the question asked provided we substitute definite integrals in place of the finite summations. This substitution gives us, for the de-

sired *a posteriori* probability that x had a value between x_1 and x_2, the formula

$$(4) \qquad p(x_1, x_2) = \frac{\int_{x_1}^{x_2} w(x) x^T (1-x)^{N-T} dx}{\int_0^1 w(x) x^T (1-x)^{N-T} dx}$$

Equation (4) is useless until the form of the *a priori existence* function $w(x)$ is specified; this depends on the *way* in which the line OS was drawn. Bayes assumed that the line OS, of unknown distance from AD, was drawn through the point of rest corresponding to a preliminary roll of the ball. This amounts to postulating that all values of x, between 0 and 1 were *a priori* equally likely. In other words, with Bayes, the *a priori existence function* $w(x)$ was a constant which, therefore, did not have to be taken into consideration.[1] Thus, instead of equation (4), Bayes gave the equivalent of the following restricted formula:

$$(5) \qquad p(x_1, x_2) = \frac{\int_{x_1}^{x_2} x^T (1-x)^{N-T} dx}{\int_0^1 x^T (1-x)^{N-T} dx}$$

I say "the equivalent of" (5) because in Bayes' day definite integrals were expressed in terms of corresponding areas.

Equation (5) constitutes Proposition 9 of the essay, but is usually referred to as Bayes' theorem.

[1] The existence function $w(x)$ does not appear either explicitly or implicitly anywhere in Bayes' essay. This fact raises the question as to whether or not Bayes had any notion of the *general* problem of causes.

IV.

Equation (5) is a very beautiful formula; but we must be cautious. More than one high authority has insinuated that its beauty is only skin deep. Speaking of Laplace's generalization and extension of the theorem, George Chrystal, the English mathematician and actuary, closed a severe attack on the whole theory of *a posteriori* probability[1] with the statement that "Practical people like the Actuaries, however much they may justly respect Laplace, should not air his weaknesses in their annual examinations. The indiscretions of great men should be quietly allowed to be forgotten."

Chrystal's advice as to the attitude one should assume toward "the indiscretions of great men" is excellent, but in the case under consideration, it was the plaintiff rather than the defendant who committed indiscretions; this is discussed in a paper by E. T. Whittaker[2] entitled "On Some Disputed Questions of Probability."

The discussions and disputes, which began shortly after the birth of the formula in 1763 and which have not as yet subsided, may be divided into two classes:

1. Discussions concerning problems in which it is known that the *a priori* existence function is not a constant.
2. Discussions concerning problems in which nothing whatever is known concerning the *a priori* existence function.

The discussions of Class 1 are out of order in so far as Bayes' theorem is concerned; recourse should be had to formula (4), Laplace's generalization of the Bayes' theorem, when it is known that $\omega(x)$ is not a constant. Failure to differentiate

[1] "On Some Fundamental Principles in the Theory of Probability," *Transactions of the Actuarial Society of Edinburgh*, Vol. 11, No. 13.

[2] *Transactions of the Faculty of Actuaries in Scotland*, Vol. VIII, Session 1919-1920.

explicitly between equations (4) and (5) has created a great deal of confusion of thought concerning the probability of causes. The discussions of Class 2 have centered on what Boole called "the equal distribution of our knowledge, or rather of our ignorance," that is to say "the assigning to different states of things of which we know nothing, and upon the very ground that we know nothing, equal degrees of probability." Regarding the legitimacy of this procedure Bayes himself contributed a very important scholium, which appeared in his essay on pages 392 and 393. The argument in this scholium, based on a corollary to Proposition 8 of the essay, may be summarized as follows:

Assuming that all values of x are *a priori* equally likely and that the N throws of a ball on the table have *not yet* been made, the probability that T times the ball will rest to the right of OS and that the remaining $N-T$ times it will rest to the left of OS is (as shown in the corollary)

$$(6) \qquad p = \int_0^1 \binom{N}{T} x^T (1-x)^{N-T} dx = \frac{1}{N+1}$$

a result in which T does not appear In other words, any assigned outcome for the throws is no more, or no less, likely than any other outcome, if *a priori* all values of x are equally likely. But, wrote Bayes in the scholium, when we say that we have no knowledge whatever *a priori* regarding the ratio x, do we not really mean that we are in the dark as to what will be the outcome when we proceed to make N throws? If so, then equation (6) justifies the assumption that *a priori* all values of x are equally likely.

To clinch his argument it must be shown that the converse of equation (6) is true That is, it must be shown that, if any outcome of throws *not yet* made is as likely as any other, then

any value of x is *a priori* as likely as any other. This converse theorem was submitted to Dr. F. H Murray, who obtained an elegant proof based on a theorem of Stieltjes.[1]

In view of Bayes' corollary and his scholium, an analysis of our bag problem with reference to the "equal distribution of our knowledge, or ignorance" is in order.

Consider again Case 1 where each drawn ball is replaced in the bag before the next drawing is made.

Assuming each of the $(M+1)$ permissible hypotheses to be *a priori* equally likely, the probability that N drawings, *not yet* made, will result in T white and $N-T$ black balls is

$$(7) \qquad P = \sum_{k=0}^{M} \frac{1}{M+1} \binom{N}{T} \left(\frac{k}{M}\right)^{T} \left(1-\frac{k}{M}\right)^{N-T}$$

Equation (7) is not, in general, independent of T[2] so that any one assigned outcome of N drawings is not as likely as any other outcome. This result is disturbing; at first sight it seems to discredit Bayes' scholium. We must, therefore, look into the the matter more closely.

Bayes' problem corresponds to drawings from a bag containing an infinite number of balls. Therefore, even if drawn balls are replaced, the chance of a particular ball being drawn more than once is zero. But when N drawings with replacements are made from a bag containing a *finite* number, M, of balls, we are by no means certain of drawing N different balls;

---

[1] *Bulletin of the American Mathematical Society*, February, 1930.

[2] Consider, for example, the case of $M = 2$. Equation (7) reduces to

$$P = \frac{1}{3}\left(\frac{1}{2}\right)^{N}\binom{N}{T}$$

a result which is not independent of T.

a particular white ball may be drawn several times over, and, likewise, a particular black ball may appear more than once. It is not surprising, therefore, that Case 1 of the bag problem does not confirm Bayes' corollary.

Consider now Case 2, where the drawn balls are not returned to the bag. If k of the total balls are white and the rest black, the probability that a sample of N balls from the bag will contain T white and $N-T$ black is

$$\binom{k}{T}\binom{M-k}{N-T}\Big/\binom{M}{N}$$

Hence, if the permissible values 0, 1, 2, 3, . . . M for k are all equally likely *a priori*, we obtain instead of (7),

$$(8) \qquad p = \sum_{k=0}^{M}\left(\frac{1}{M+1}\right)\binom{k}{T}\binom{M-k}{N-T}\Big/\binom{M}{N} = \frac{1}{N+1}$$

a result independent of any assigned value for T and identical with the result in the corollary to Proposition 8 of the essay.

SUMMARY

Bayes' theorem is the answer to a special case of the general problem of causes. The special case postulates that the *a priori* existence probabilities for the various admissible causes of an observed event are equal.

In the essay Bayes recommends that his theorem be adopted whenever we find ourselves confronted with total ignorance as to which one of several possible causes produced an observed event. To justify this recommendation Bayes takes the attitude that: A state of total ignorance regarding the causes of an ob-

served event is equivalent to the same state of total ignorance as to what the result will be if the trial or experiment has not yet been made. This interpretation is a generalization of the fact that in his billiard table problem, the assumption of equal likelihood for all possible positions of the line OS, gives equal probabilities for the various possible outcomes of a set of N ball rollings not yet made.

Laplace, Poincaré and Edgeworth[1] have shown that the *a priori existence* function $w(x)$, which appears in the Laplacian generalization of Bayes' theorem, is of negligible importance when the numbers N and T are large Therefore, when this condition holds, one need not hesitate to use Bayes' restricted formula for the solution of a problem of causes

The transmission, by Price, of Bayes' posthumous essay to the Royal Society marked an epoch in the history of the literature on probability theory. As mentioned at the beginning of this paper, Karl Pearson has called the extension of Bayes' problem the "Fundamental Problem of Practical Statistics."

[1] Laplace: "Oeuvres," Vol. 9, p. 470 Poincaré. "Calcul des Probabilités," 2d edition, p. 255. Bowley: "F. Y. Edgeworth's Contribution to Mathematical Statistics," pp. 11 and 12.

E. C. Molina

ON CERTAIN PROPERTIES OF FREQUENCY DISTRIBUTIONS OBTAINED BY A LINEAR FRACTIONAL TRANSFORMATION OF THE VARIATES OF A GIVEN DISTRIBUTION

By

H. L. Rietz

Considerable evidence has been presented by R. A. Fisher[1] to show that, by an appropriate transformation $z = f(r)$ of small sample correlation coefficients $r\,(-1 \leqq r \leqq 1)$ distributed in accord with a decidedly skew frequency curve, values of z are obtained which are distributed nearly in a normal distribution. In fact, the approach of the distribution of z to normality seems sufficiently rapid to justify the use of the probable error of z in many applications as if it were normally distributed. Such a change in the character of the distribution of an important statistic suggests the further study of properties of the distribution of variables obtained by applying rather simple transformations to variates distributed from -1 to $+1$ in accord with a given frequency function. In a previous paper,[2] the writer has dealt with a similar problem when each variate of a given unimodal distribution of any finite range is replaced by a given power of the variate.

Consider a positive unimodal continuous frequency function

[1] Metron, Vol. 1, Part 4 (1921) pp 3-32

[2] Proceedings of the National Academy, Vol. 13, No. 12 (1927), pp. 817-820.

$y = \psi(x)$ of a system of variates $x_1, x_2, \ldots x_n$ with a range of -1 to $+1$, with $\psi(-1) = \psi(1) = 0$, with a single mode at some point, say at $x = b$ $(-1 < b < 1)$, and with the derivative $\psi'(x)$ continuous More precisely, we assume that $\psi(x)$ is positive except at the end points of the interval -1 to $+1$, where it is zero, and that $\psi'(x)$ changes from positive to negative at $x = b$, and is non-negative or non-positive at any point $x = a$ according as a is less or greater than b.

It is the main object of the present paper to consider certain properties of the distribution of variates $u_i = (ex_i + f)/(gx_i + h)$ obtained by a linear fractional transformation of the x's, where e, f, g, and h are real numbers so selected that $u = (ex + f)/(gx + h)$ is continuous from $x = -1$ to $x = 1$.

When $g = 0$, we have the case of the linear transformation which simply has an effect equivalent to a change of origin and of unit of measurement. As we are not in the present problem much interested in such a simple transformation, we shall, in general, assume $g \neq 0$. Moreover, we take g positive, since this involves no loss of generality.

We shall, except as otherwise stated, restrict our considerations to the interval for u that corresponds to $-1 \leq x \leq 1$, and to such transformations that the derivative of u with respect to x is finite for each value of x and that u increases when x increases These restrictions require that

$$\frac{du}{dx} = \frac{he - fg}{(gx + h)^2}$$

where $g < |h|$ and where the determinant

$$(1) \qquad he - fg = \begin{vmatrix} e & f \\ g & h \end{vmatrix} > 0$$

Starting then with

(2) $$u = \frac{ex+f}{gx+h},$$

we have

(3) $$x = \frac{f-hu}{gu-e}$$

Next, let

(4) $$v = \phi(u)$$

be the frequency function of the new variates u. Then we may write[1]

(5) $$v = \phi(u) = \psi\left(\frac{f-hu}{gu-e}\right) \cdot \frac{he-fg}{(gu-e)^2}.$$

Since $he-fg > 0$, we know that v is positive throughout the interval in which we are interested except that $V=0$ at the end points. From (5) it seems that the new distribution function may possibly become infinite when $u = e/g$, but the question then arises as to whether e/g is an admissible value of u.

We shall prove that e/g is not an admissible value of u by showing that u cannot take the value e/g within the interval $u=(f-e)/(h-g)$ to $u=(e+f)/(g+h)$ wherein u lies when $-1 \leqq x \leqq 1$. In this connection we shall also establish some inequalities that will be found useful in the consideration of certain properties of the new distribution. Consider first the cases in which $g+h$ is positive.

Then since $eh > fg$, we have $eh+eg > fg+eg$.

Divide by $g(g+h)$, and we have $\frac{e}{g} > \frac{f+e}{g+h}$ Hence,

[1] cf. Annals of Mathematics, vol. 23, No. 4 (1922), pp. 293-4.

e/g is too large when $g+h$ is positive to be an admissible value of u.

Consider next the cases in which $g+h$ is negative. In this case, $h<0$ since $g>0$. Hence $g-h>0$. Then since $eh>fg$, we have $eh-eg>fg-eg$. Divide by the positive number $g(g-h)$ This gives $\frac{-e}{g}>\frac{f-e}{g-h}$ and $\frac{e}{g}<\frac{e-f}{g-h}$.

Hence, when $g(g+h)<0$, e/g is too small to be an admissible value of u.

To summarize with $g>0$, we have shown that:

(a) When $g+h$ is positive, e/g is too large to be an admissible value of u

(b) When $g+h$ is negative, e/g is too small to be an admissible value of u.

Returning now to the consideration of our frequency function $y=\psi\left(\frac{f-hu}{gu-e}\right)\cdot\frac{he-fg}{(gu-e)^2}$ in (5), we obtain

$$(6)\quad \frac{dy}{du}=\frac{(he-fg)^2}{(gu-e)^4}\,\psi'\left(\frac{f-hu}{gu-e}\right)-\frac{2g(he-fg)}{(gu-e)^3}\,\psi\left(\frac{f-hu}{gu-e}\right).$$

When u takes the value $(eb+f)/(gb+h)$ into which variates at the mode $x=b$ are transformed, we know that $\psi'\left(\frac{f-hu}{gu-e}\right)=\psi'(b)=0$.

By making use of the fact that $he-fg>0$, and the propositions (a) and (b) relating to the inadmissibility of e/g as a value of u in an examination of the right hand member of (6) for $u=(eb+f)/(gb+h)$ we establish the following proposition in regard to the sign of the derivative dy/du for the value of u which corresponds to the modal value of x

When $g+h\neq 0$, dy/du is positive or negative at $u=(eb+f)/(gb+h)$ according as $g+h$ is **positive or negative.**

The truth of this proposition follows readily by applying (a) and (b) to (6), remembering that g is positive and that $\psi'(b)$ vanishes.

We shall show next in case $g+h > 0$, that dv/du is non-negative for all admissible values of u less than $(eb+f)/(gb+h)$. To see this from (6), note first that $\psi'[(f-hu)/(gu-e)]$ remains non-negative for $(f-hu)/(gu-e) < b$ or for u less than $(eb+f)/(gb+h)$, and note second that $g/(gu-e)^3$ is negative since e/g is too large to be an admissible value of u under the condition $g+h > 0$

Next, in case $g+h < 0$, dv/du is non-positive for all values of $u > (eb+f)/(gb+h)$. To see this from (6), note first that $\psi'[(f-hu)/(gu-e)]$ remains non-positive for $(f-hu)/(gu-e) > b$ or for $u > \frac{(eb+f)}{(gb+h)}$, and note second that $g/(gu-e)^3$ is positive when $g+h < 0$ because in this case $u > e/g$.

To summarize, when $g+h \neq 0$, we state the

Theorem I. *When the derivative dv/du is positive for the value of u into which variates at the modal value $x=b$ transform, then dv/du is non-negative for all smaller values of u. Similarly, when dv/du is negative for the value of u into which variates at the modal value $x=b$ transform, then dv/du is non-positive for all larger values of u.*

Finally, we wish to inquire about a modal value for the frequency function $v=\phi(u)$ in (5). To this end, consider first the case in which dv/du is positive at $u=(eb+f)/(gb+h)$. At a point between $u=(eb+f)/(gb+h)$ and the upper bound of u, that is $(e+f)/(g+h)$, a maximum value of v occurs. To

see this, note when $u = (e+f)/(g+h)$ that $dv/du = \psi'(1)(g+h)^4/(he-fg)^2$ which is negative, or zero since $\psi'(1)$ is negative or zero. If it is negative, there is a maximum where the sign of the continuous first derivative changes from positive to negative If dv/du is zero at $u = (e+f)/(g+h)$, it follows also that there is at least one maximum of $v = \phi(u)$ between $u = (eb+f)/(gb+h)$ and $u = (e+f)/(g+h)$ since $v = 0$ at $u = (e+f)/(g+h)$ and v must have changed from an increasing positive function at $u = (eb+f)/(gb+h)$ to a decreasing function before becoming zero at $u = (e+f)/(g+h)$. Similarly, it may be shown that there is a mode at a value of $u < (eb+f)/(gb+h)$ whenever dv/du is negative at $u = (eb+f)/(gb+h)$.

We may then state the following:

Theorem II. *Given a unimodal continuous positive function* $y = \psi(x)$ *of variates* x, *with a range from* -1 *to* $+1$, *with a mode at* $x = b\ (-1 < b < 1)$, *with* $\psi(-1) = \psi(1) = 0$, *and with the derivative* $\psi' x$ *continuous from* $x = -1$ *to* $x = 1$, *then the frequency distribution* $v = \phi(u)$ *of variates* $u = (ex+f)/(gx+h)$ $(g > 0)$ *has a mode at a value of* $u > (eb+f)/(gb+h)$ *when* $g+h > 0$. *It has a mode at a value of* $u < (eb+f)/(gb+h)$ *when* $g+h < 0$.

Since we have so restricted our transformation $u = \frac{(ex+f)}{(gx+h)}$ that the order of corresponding values is preserved, the transformation carries the median of the distribution of x's into the median of the distribution of u's, and we may state the following:

Corollary. *If* $y = \psi(x)$ *has its median and mode coincident at* $x = b$, *the frequency distribution* $v = \phi(u)$ *of* $u = (ex+f)/(gx+h)$ *has a modal value greater or less than its median according as* $g+h$ *is greater or less than zero.*

Thus far we have imposed the condition $g < |h|$. Let us next consider the cases in which $h = -g$ and $h = g$ instead of requiring that $g < |h|$. Consider first the case $h = -g$. In this case

$$(7) \qquad u = \frac{1}{g} \cdot \frac{ex+f}{x-1}$$

and

$$(8) \qquad \frac{du}{dx} = \frac{he-fg}{(gx+h)^2} = -\frac{e+f}{g(x-1)^2}.$$

Both u and du/dx become infinite as x approaches 1. Suppose e and f so chosen that u is an increasing function of x for the interval $-1 \leqq x < 1$, then u in (7) is an increasing function of x for the larger interval $-\infty < x < 1$; and it follows, for the case $h = -g$, that e/g is too small to be an admissible value of u when $-1 \leq x < 1$, since it is the value of u when $x = -\infty$.

For the case $h = g$, we have

$$(9) \qquad u = \frac{ex+f}{g(x+1)}$$

and

$$(10) \qquad \frac{du}{dx} = \frac{e-f}{g(x+1)^2}.$$

Since u in (9) is an increasing continuous function of x for the interval $-1 < x < \infty$ wherever e and f are so selected that it is increasing for the sub-interval $-1 < x \leqq 1$, it follows, for $h = g$, that e/g, the value of u when $x = \infty$, is too large to be an admissible value of u when $-1 < x \leqq 1$. By making use of the fact that e/g is too small or too large

to be an admissible value of u according as $h=-g$ or $+g$, we readily obtain the following results from an examination of (6): The derivative dv/du given in (6) is positive at the point $u=(eb+f)/(gb+h)$ when $h=g$, and it is negative at this point when $h=-g$

Moreover it readily follows as in the case where $g<|h|$ that when the derivative dv/du is positive for the value of u into which the modal $x=b$ transforms, then dv/du is non-negative for all smaller values of u, and when dv/du is negative for the value of u into which the modal value $x=b$ transforms, it is non-positive for all larger values of u

Next, for the case $h=g$, a mode occurs for a value of $u>(eb+f)/(gb+h)$. This may be seen by noting that as x approaches 1 and as u takes corresponding values dv/du in (6) approaches the value $16\, g^2\, \psi'(1)/(e-f)^2$ which is negative or zero. The analysis given above for the corresponding case $g<|h|$ may be applied, with the conclusions stated in Theroem II by replacing $g+h>0$ by $h=g$ and $g+h<0$ by $h=-g$

The question very naturally arises as to whether there exists a linear fractional transformation $u=(ex+f)/(gx+h)$ that will transform almost any distribution with the properties of $y=\psi(x)$ into a new distribution $v=\phi(u)$ with a mode at a previously assigned point $u=c$ within the range of admissible values of u. To insure a mode for $v=\phi(u)$ at $u=c$, it is, of course, sufficient that there exist values of e, f, g, and h that make the continuous function

$$(11)\quad \frac{dv}{du}=\frac{(he-fg)^2}{(gu-e)^4}\,\psi'\left(\frac{f-hu}{gu-e}\right)-\frac{2g(he-fg)}{(gu-e)^3}\,\psi\left(\frac{f-hu}{gu-e}\right)$$

change sign from positive to negative at $u=c$

Since the only restrictions on e, f, g, and h are that

they shall be real, and that g and $he-fg$ shall be positive, it seems that the requirement that dv/du shall change from positive to negative at an assigned value of u could probably be satisfied for some important classes of relatively simple functions. As a simple example, take the quadratic function $\psi(x) = Ax^2 + Bx + C$, which, when subjected to the conditions on $\psi(x)$, becomes $\psi(x) = 3(1-x^2)/4$.

The mode is in this case at $x = 0$. The problem we propose is to find the linear fractional transformation $u = (ex+f)/(gx+h)$ that will transform $\psi(x)$ into $v = \phi(u)$ with a mode at an assigned $u = c$ In this case (11) becomes

$$\frac{dv}{du} = -\frac{3}{2}\frac{he-fg}{(gu-e)^5}\Big\{(he-fg)(f-hu) - \tag{12}$$
$$g\left[(g^2-h^2)u^2 + 2u(fh-eg) + e^2 - f^2\right]\Big\}.$$

To facilitate the examination of (12), make $h = g$. Then (12) reduces to

$$\frac{dv}{du} = -\frac{3}{2}\,\frac{g^2(e-f)^2}{(gu-e)^5}\,(e+2f-3gu). \tag{13}$$

Since $g+h > 0$, we have $gu-e < 0$, and consequently the coefficient of $(e+2f-3gu)$ is positive To provide for the change of sign of (13) at $u = c$, select e, f, and g so that $e+2f = 3cg$. To make (13) positive at $u = c-\delta$ and negative at $u = c+\delta$, where δ is arbitrarily small and positive, we may assign to g any positive value and to e any value greater than cg, for then f is less than e, which is the condition $he-fg > 0$ when $h = g$ While there are thus an infinite number of ways in which we may select a linear

fractional transformation so that, when applied to special functions, it will give a new distribution with a mode at an assigned point, no general proposition is proved that assures an assigned modal value of $\psi(x)$.

H. L. Rietz

ON SMALL SAMPLES FROM CERTAIN NON-NORMAL UNIVERSES*

By

PAUL R RIDER
Washington University

INTRODUCTION

The distribution of the ratio

$$z = \frac{\text{mean of sample} - \text{mean of universe}}{\text{standard deviation of sample}}$$

which is of great importance in the theory of small samples, has been derived exactly by theoretical methods for samples of any size from a normal universe.[1] Experimental studies[2] have been

*The writer desires to express his grateful appreciation to the National Research Council, which made possible this study by a grant-in-aid for the assistance of a computer.

[1] See, for example, R. A Fisher, Applications of "Student's" Distribution, Metron, vol 5, No. 3 (Dec. 1, 1925), pp. 90-104.

[2] e. g. W. A Shewhart and F. W. Winters, Small Samples—New Experimental Results, Journal of the American Statistical Association, Vol 23 (1928), pp. 144-53;

J. Neyman and E. S Pearson, On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference Part I, Biometrika, Vol 20A (1928), pp. 175-240;

"Sophister," Discussion of Small Samples Drawn from an Infinite Skew Population, Biometrika, Vol. 20A (1928), pp. 389-423,

E S Pearson assisted by N. K. Adyanthāya and others, The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations 2nd paper, Biometrika, Vol 21 (1929), pp. 259-86.

made of the z -distribution for samples of specific sizes from other types of universe A theoretical method applicable to samples from a discrete universe was used in a previous paper,[1] in which a rectangular universe was studied in some detail. The rectangular universe was chosen as being the simplest from the standpoint of the method employed, and as a good example of a limited symmetric distribution It is the purpose of the present paper to apply the method to a triangular population, which is a specimen of a limited skew distribution, and also to a U-shaped universe. The rectangular, triangular and U-shaped universes are shown in Table I in the columns headed R , T , and U , respectively. Their graphs are exhibited in Figure 1.

In addition to the z -distribution, the distributions of means from the triangular and from the U-shaped universe are given.

In the concluding section is discussed the probability corresponding to an interval of three sample standard deviations on each side of the sample mean.

All of the results of the paper are for samples of four.

THE DISTRIBUTION OF z

The distributions of z are shown in Table II,[2], in which the distribution for samples from a normal universe, N , is also given.

The cumulated probability of z for the triangular and for the U-shaped universe are shown in Table III, which may be compared with a similar table for a rectangular and for a normal universe given in Biometrika, Vol. 21 (1929), p. 131.

[1] P R. Rider, On the Distribution of the Ratio of Mean to Standard Deviation in Small Samples from Non-normal Universes, Biometrika, Vol 21 (1929), pp. 124-143.

[2] For an explanation of the method of deriving these distributions see Rider, loc. cit.

These cumulated probabilities are plotted on probability paper in Figures 2 and 3 and may be compared with similar probabilities for a rectangular universe by reference to Biometrika, Vol. 21 (1929), p. 129, Figure 2.

The principal results to be noted are as follows:

1. The general characteristics of the z-distribution for the U-shaped universe are the same as those for a rectangular universe, viz. a greater number of z's outside of a certain value of $|z|$, and also a greater clustering of z's about the origin, than is the case for a normal universe.[1] This is to be expected, since the values of β_2 for U and R are 1.132 and 1.776 respectively, as compared with the value 3 for N.

2. The negative skewness in the triangular universe produces skewness of the opposite type in the distribution of z, as found experimentally by Neyman and E. S. Pearson[2] and by "Sophister."[3] This means (in the case of negative skewness in the universe) that the probability corresponding to an interval from $-\infty$ to z is smaller than when the sampling is from a normal universe.

3 The cumulated probability of $|z|$, or the probability corresponding to an interval from $-z$ to z, is somewhat the same for the triangular universe as for a normal universe;[4] a comparison is made in Table IV.

Results 2 and 3 are apparently due to the fact that in a

---

[1] See Rider, loc. cit., p. 130.

[2] Biometrika, Vol. 20A (1928), p 198

[3] Biometrika, Vol. 20A (1928), p. 408.

[4] cf. E. S. Pearson assisted by N. K. Adyanthāya and others. The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations. 2nd paper, Biometrika, Vol. 21 (1929), pp. 259-86

skew universe the regression of variance on mean[1] is often essentially linear (if parabolic, the vertex of the parabola is well to one side of the scatter diagram). Let us consider the case in which the slope of the regression line is positive Designating by x the difference between the mean of a sample and the mean of the universe, and by s the standard deviation of the sample, we see that large values of $|x|$ tend to be associated with large values of s^2 (and therefore with large values of s). Thus the values of z tend to be smaller. On the other hand, for large values of $|-x|$, s is smaller and $|-z|$ consequently larger. This means that the frequencies corresponding to the algebraically lower values of z are greater than in the case of a normal universe, or that the use of "Student's" tables would give results too small for the probability that the mean of a sample does not exceed *algebraically* the mean of the universe by more than z times the standard deviation of the sample. The opposite is true in the case studied here, since the universe is negatively skew and the regression line of s^2 on x would have a negative slope.

Since there is a shifting of the whole cumulated z-distribution to the right or left, the effect noted in 3 is readily explained. As a result of this effect we should apparently not be far wrong, when sampling from a skew universe, if we used "Student's" tables to obtain the probability that the mean of a sample does not exceed *numerically* the mean of the universe by more than z times the standard deviation of the sample.[2]

---

[1] For the regression formula see J Neyman, On the Correlation of the Mean and the Variance in Samples from an "Infinite" Population, Biometrika, Vol 18 (1926), pp 401-13.

[2] See E S. Pearson assisted by N. K. Adyanthāya and others. The Distribution of [illegible] ial Symmetrical [illegible] 2 (1929), pp 259-8[illegible]

THE DISTRIBUTION OF MEANS OF SAMPLES

The distributions of means of samples are shown in Tables V and VI. In these tables x indicates the difference between the mean of the sample and the mean of the universe.

For the difficulties involved in obtaining satisfactory results for the distribution of means of small samples from a U-shaped universe see K. J Holtzinger and A. E. R. Church, "On the Means of Samples from a U-shaped Population," Biometrika, Vol 20A (1928), pp 361-88.

The probability corresponding to an interval of three sample standard deviations on each side of the sample mean.

If M is the mean and σ the standard deviation of a normally distributed variate X, then, as is well known, the probability that an item selected at random will lie within the range $M \pm 3\sigma$ is 0.997. If $\bar{X}$ and s are the mean and the standard deviation respectively of a sample, the expected or average probability corresponding to the interval $\bar{X} \pm 3s$ will be different from the probability corresponding to the interval $M \pm 3\sigma$. Shewhart[1] obtained experimentally for the average probability for samples of four associated with the interval $\bar{X} \pm 3s$ the values 0.90 for a normal universe, 0.91 for a rectangular universe, and 0.91 for a triangular universe.

By analyzing all possible samples of four from the rectangular and triangular universes of Table I it was possible to obtain the probability corresponding to an interval of $3s$ on either side of the sample mean. For example let us consider the sample (1, 1, 2, 2), for which $\bar{X} = 1.5$, $s = 0.5$. The interval $\bar{X} \pm 3s$ extends from 0 to 3. This interval includes 0.4 of the rectangular universe R; 0.4 then is the probability that an

[1] W A. Shewhart, Note on the Probability Associated with the Error of a Single Observation, Journal of Forestry, Vol 26 (1928) pp. 601-607

observed value will fall within the interval Now the particular sample (1, 1, 2, 2) would occur 6 times out of 10,000. If we take all of the samples for which the interval $X \pm 3s$ includes 0.4 of the rectangular universe we find that such samples occur 106 times out of 10,000. Such an analysis leads to Table VII, from which it is ascertained that the average probability corresponding to an interval of $X \pm 3s$ is 0 920. A similar analysis of the triangular universe T gives us Table VIII and yields 0.907 as the average probability associated with $X \pm 3s$. A better understanding of the situation may be obtained from Figure 4.

Paul R. Rider

TABLE I

Rectangular, Triangular and U-Shaped Universes

X	FREQUENCY		
	R	*T*	*U*
0	1		10
1	1	1	5
2	1	2	1
3	1	3	1
4	1	4	1
5	1	5	1
6	1	6	1
7	1	7	1
8	1	8	5
9	1	9	10
10		10	
Total	10	55	36
Mean	4.5	7	4.5
β_1*	0	0 326	0
β_2*	$1.\dot{7}\dot{7}5$	$2.3\dot{6}$	1.132+

*The values of the β's are uncorrected for grouping. The dots over the digits indicate repeating decimals. The values for a continuous rectangular distribution are $\beta_1 = 0$, $\beta_2 = 1.8$, and for a continuous triangular distribution are $\beta_1 = 0.32$, $\beta_2 = 2.4$.

TABLE II

Probability of z for Samples of 4

z	N	R	T	U
Below −4.25	0026	.0077	.0015 +	.0384
−4.25 to −3.75	.0011	.0022	0012	0004
−3.75 to −3.25	.0018	.0026	.0007	.0009
−3 25 to −2.75	.0032	0032	.0032	.0077
−2.75 to −2.25	0062	0074	.0028	0016
−2.25 to −1 75	.0131	0188	.0061	.0106
−1.75 to −1.25	.0314	.0267	.0251	.0147
−1.25 to −0 75	.0829	0692	.0615	.0256
−0.75 to −0.25	.2047	.2000	.2098	.2299
−0.25 to 0.25	.3058	.3244	.3249	.3405 +
0.25 to 0 75	.2047	.2000	1741	.2299
0.75 to 1 25	0829	0692	.0764	.0256
1.25 to 1.75	.0314	.0267	0566	.0147
1 75 to 2.25	.0131	0188	.0118	.0106
2 25 to 2.75	.0062	.0074	0094	0016
2.75 to 3.25	.0032	.0032	.0174	.0077
3.25 to 3.75	.0018	.0026	.0000	.0009
3.75 to 4.25	.0011	.0022	.0025 +	.0004
Above 4.25	.0026	.0077	.0150 −	.0383

TABLE III

The cumulated probability of z , or probability that the mean of a random sample of 4 will not exceed (in algebraic sense) the mean of the universe by more than z times the standard deviation of the sample

z	Cumulated Probability Triangular Universe		Cumulated Probability U-Shaped Universe	
	for - z	for z	for - z	for z
0.0	51955−	.51955−	54355+	54355+
.1	.41649	54037	39365−	.60635+
.2	.34497	.61053	.34651	.65349
.3	.28885+	.65136	.30739	.69261
4	.22719	.70010	.27831	.72193
.5	18568	74269	.22081	.77991
.6	.14350−	.76942	.14785+	.85215−
.7	.11580	79993	.11382	88618
.8	.09485−	81086	09844	90192
9	.07784	.83462	.09065+	.90935+
1 0	.06130	.86748	.08285−	.91715+
1.1	05053	87456	07994	.92006
1 2	.04256	88731	.07471	.92529
1.3	03716	88731	07363	.92637
1.4	.03152	.90787	.07179	.92821
1.5	.02783	.91316	.06614	.93387
1 6	02334	.91911	05979	94021
1 7	01845−	93480	.05975−	.94025−
1 8	.01552	.94390	.05941	.94059
1.9	.01410	.94390	.05798	94202
2 0	.01366	94810	.05441	.94774
2 1	.01265−	94810	.04959	.95041
2 2	.01039	.95565−	.04892	.95108
2 3	.00907	95565−	.04892	.95108
2.4	.00871	95565−	.04891	.95109
2.5	.00816	.95565−	04891	.95118
2 6	.00725+	.95565−	.04803	.95197
2.7	.00725+	95565−	.04732	.95268
2 8	.00661	96509	.04728	.95272
2.9	00483	97910	.04133	.95867
3.0	.00462	98250−	.03954	.96046
3 5	.00272	98250−	.03904	96132
4.0	00242	98250−	.03833	.96168

TABLE IV

Cumulated Probability of $|z|$ for Samples of 4.

$\|z\|$ greater than	Probability: Triangular Universe	Probability: Normal Universe	$\|z\|$ greater than	Probability: Triangular Universe	Probability: Normal Universe
0.0	.9219	1.0000	1.6	1042	.0695−
1	.8761	8735+	1 7	0836	.0603
.2	.7303	7519	1 8	0716	.0526
.3	6375−	.6392	1 9	.0702	.0460
.4	.5271	5382	2 0	.0652	.0405+
.5	.4423	.4502	2 1	.0646	.0358
.6	3723	3751	2.2	0547	.0318
.7	3135−	3121	2.3	.0534	.0283
8	.2834	2599	2 4	0531	.0253
9	2432	.2169	2 5	.0525+	0227
1 0	.1891	.1817	2.6	0516	.0204
1 1	.1755−	.1528	2.7	.0516	.0185−
1.2	.1552	.1292	·2 8	.0415+	0167
1 3	.1497	.1098	2 9	.0257	.0152
1 4	.1236	.0938	3.0	.0212	.0138
1.5	.1146	0805+			

TABLE V

Distribution of Means of Samples of 4 from Triangular Universe

x	Probability	x	Probability	x	Probability
-5.25	.00001	-2.25	.01627	0.75	.07202
-5.00	.00004	-2.00	.02200	1.00	.06437
-4.75	.00009	-1.75	.02882	1.25	.05496
-4.50	.00019	-1.50	.03559	1.50	.04462
-4.25	.00038	-1.25	.04501	1.75	.03415+
-4.00	.00070	-1.00	.05362	2.00	.02430
-3.75	.00125	-0.75	.06187	2.25	.01569
-3.50	.00212	-0.50	.06916	2.50	.00881
-3.25	.00344	-0.25	.07484	2.75	.00393
-3.00	.00537	0.00	.07834	3.00	.00109
-2.75	.00805-	0.25	.07918		
-2.50	.01165-	0.50	.07707		

x = (mean of sample) − (mean of universe)

TABLE VI

Distribution of Means of Samples of 4 from U-Shaped Universe

x	Fre-quency	Prob-ability	x	Fre-quency	Prob-ability
-4.50	10000	.0060	0.25	106660	.0635+
-4.25	20000	0119	0.50	62755	.0374
-4 00	19000	.0113	0 75	51244	.0305+
-3.75	15000	0089	1 00	49270	.0293
-3.50	14225	0085-	1.25	48376	.0288
-3 25	15300	.0091	1.50	49505	.0295-
-3 00	16690	.0099	1 75	63960	.0381
-2.75	18140	.0108	2.00	89660	.0534
-2.50	35651	.0212	2.25	81224	.0484
-2.25	81224	.0484	2.50	35651	.0212
-2 00	89660	.0534	2.75	18140	.0108
-1.75	63960	0381	3.00	16690	.0099
-1 50	49505	.0295-	3.25	15300	0091
-1 25	48376	.0288	3.50	14225	.0085-
-1.00	49270	.0293	3.75	15000	0089
-0 75	51244	.0305+	4.00	19000	.0113
-0 50	62755	.0374	4.25	20000	.0119
-0.25	106660	.0635+	4.50	10000	.0060
0.00	146296	.0871			
			Total	1679616	1.0001

x = (mean of sample) – (mean of universe)

TABLE VII

Probability Corresponding to the Interval $\bar{X} \pm 3s$
Rectangular Universe

Proportion of universe included in $\bar{X} \pm 3s$*	Number of samples for which this proportion occurs**
0.1	10
0.2	8
0.3	84
0.4	106
0.5	284
0.6	324
0.7	564
0.8	652
0.9	888
1.0	7080
Total	10000

* i. e the probability corresponding to $\bar{X} \pm 3s$

** The probability of the occurrence of this proportion is, of course, obtained by dividing by 10000

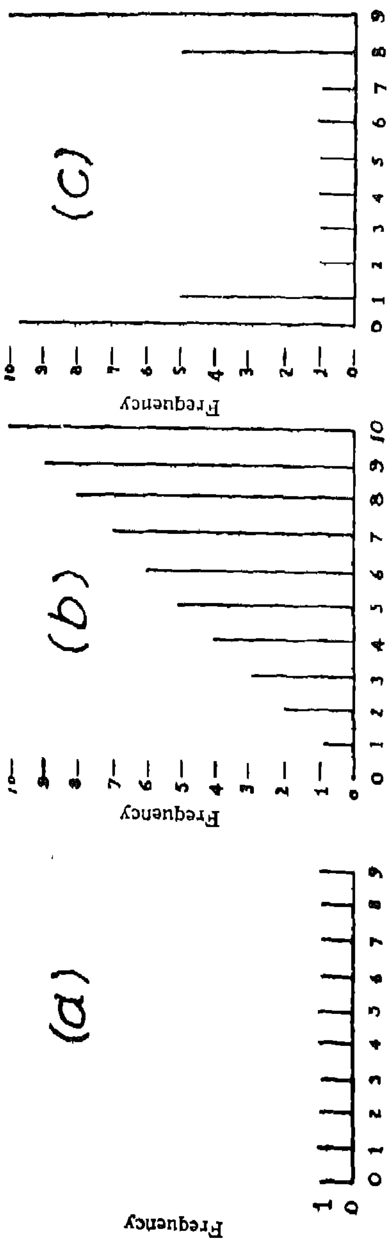
TABLE VIII

Probability Corresponding to the Interval $\bar{X} \pm 3s$

Triangular Universe

Proportion of universe included in $\bar{X} \pm 3s$	Number of samples for which this proportion occurs	Probability of occurrence of this proportion	Cumulated probability
1/55 = 018	1	—	—
2/55 = .036	16	—	—
3/55 = .055–	89	—	—
4/55 = .073	256	—	—
5/55 = .091	625	.0001	.0001
6/55 = .109	1448	.0002	.0003
7/55 = .127	2401	.0003	.0006
8/55 = .145–	4096	0004	.0010
9/55 = .164	6993	.0008	.0018
10/55 = .182	11388	.0012	0030
12/55 = .218	1280	0001	.0031
13/55 = 236	7776	.0008	.0039
14/55 = .255–	2928	.0003	.0042
15/55 = .273	8762	.0010	.0052
18/55 = .327	12768	.0014	.0066
19/55 = .345+	36000	.0039	.0105
20/55 = .364	8640	.0009	.0114
21/55 = .382	26508	.0029	.0143
22/55 = .400	5400	.0006	.0149
24/55 = .436	32768	.0036	.0185
25/55 = .455–	21600	0024	.0209
26/55 = 473	10584	.0012	.0221
27/55 = .491	112764	.0123	.0344
28/55 = .509	19698	.0022	0366
30/55 = .545+	71526	.0078	.0444
33/55 = 600	27116	.0030	.0474
34/55 = 618	296384	.0324	0798
35/55 = .636	115128	0126	.0924
36/55 = .655–	37892	.0041	.0965
39/55 = .709	54092	0059	.1024
40/55 = .727	555924	.0608	1632
42/55 = .764	57888	.0063	.1695
44/55 = .800	26416	0029	.1724
45/55 = .818	556520	.0608	.2332
49/55 = .891	,774320	.0846	.3178
52/55 = 945+	904676	,0989	.4167
54/55 = .982	879564	0961	.5128
55/55 = 1.000	4458390	4872	1 0000
Total	9150625	1 0000	

i. e. the probability corresponding to $\bar{X} \pm 3s$



Rectangular Universe

Triangular Universe

U-Shaped Universe

FIGURE 1

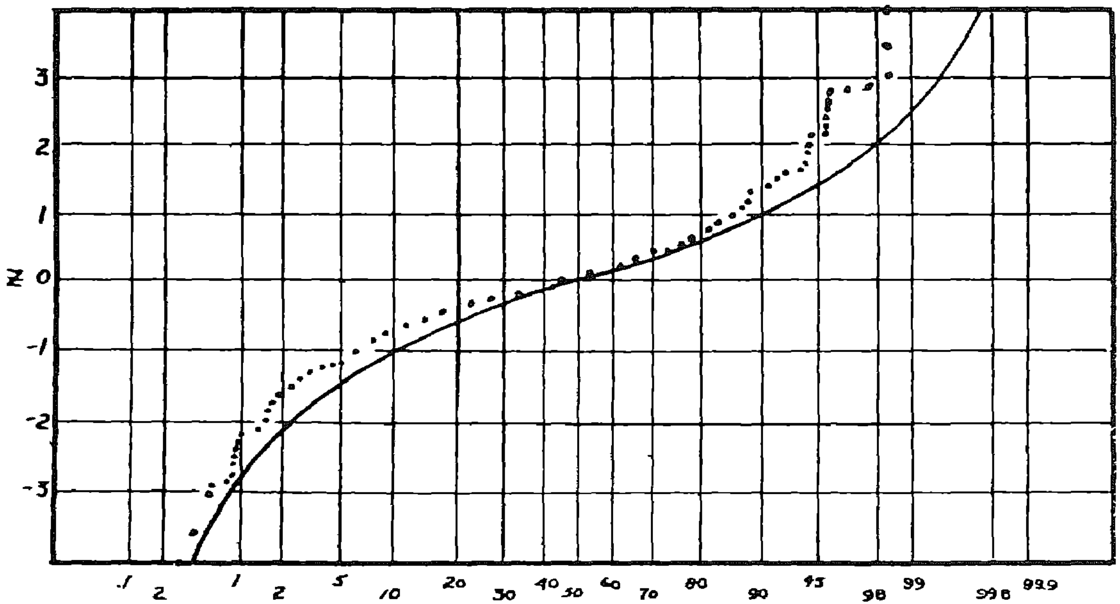


FIGURE 2

Cumulated Probability of z —Triangular Universe
The curve is for samples of 4 from a normal universe.
The dots are for samples of 4 from the universe T

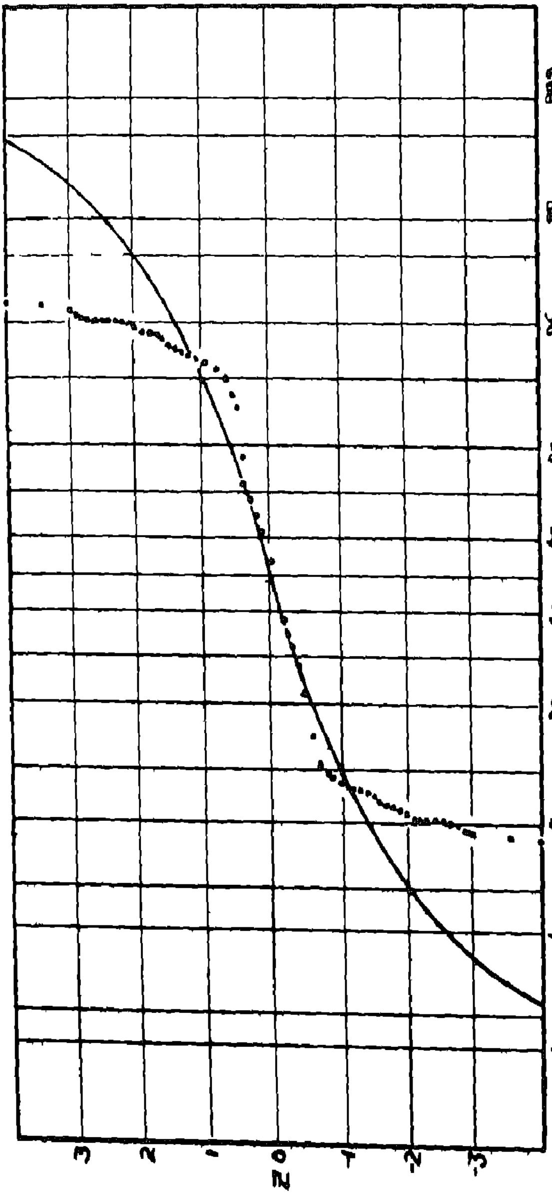


FIGURE 3

Cumulated Probability of z —U-Shaped Universe
The curve is for samples of 4 from a normal universe.
The dots are for samples of 4 from the universe U.

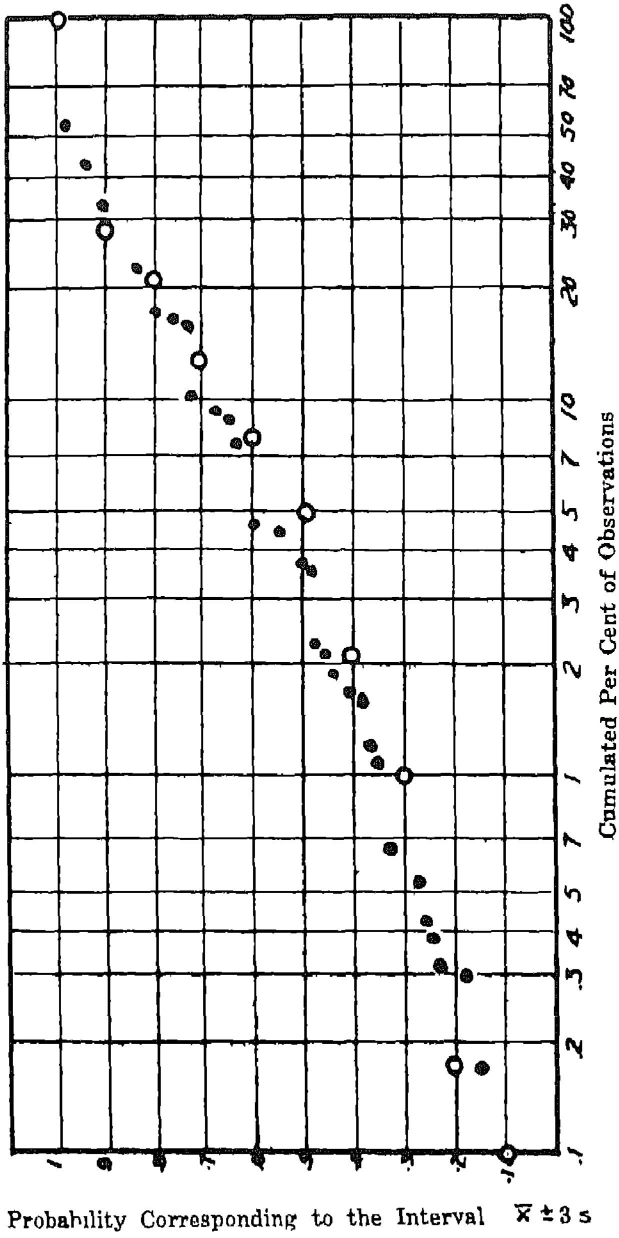


FIGURE 4

Probability Corresponding to the Interval $\bar{X} \pm 3s$
The circles are for samples of 4 from a rectangular universe,
the dots for samples of 4 from a triangular universe

AN EMPIRICAL DETERMINATION OF THE DISTRIBUTION OF MEANS, STANDARD DEVIATIONS AND CORRELATION COEFFICIENTS DRAWN FROM RECTANGULAR POPULATIONS*

By

HILDA FROST DUNLAP

Territorial Normal and Training School, Honolulu, Hawaii

Formulae for the standard errors of means, standard deviations and correlation coefficients have been derived on the assumption of a normal distribution in the sampled population. They are said to serve approximately even when the population varies considerably from the normal. This paper presents empirical evidence of their applicability in the case of means and standard deviations of samples of ten from a rectangular discontinuous population, and of correlation coefficients of samples of fifty-two from a rank distribution.

The data for the study of the distribution of means and standard deviations were secured by throwing ten dice 1600 times. The dice were cubes four-tenths of an inch along an edge and numbered on opposite faces 1-6, 2-5, 3-4. They were constructed of bone and formed a matched set.

*The writer is indebted to Jack W. Dunlap for reading the entire manuscript and for checking the mechanical computations.

These were thrown from a cup whose inside diameter was 1.75 inches and whose depth was 2.5 inches The dice were shaken in a box and then cast upon an especially prepared flat topped table covered with eight thicknesses of an army blanket.

As a guard against any possible bias in the table, the dice were thrown alternately with the right and left hands. After each throw the number of aces, deuces, treys, fours, fives, and sixes were recorded, and the mean and standard deviation calculated In this study each throw was taken as a sample of ten drawn from a population of 16,000

The next step was to determine whether there was any systematic bias in the dice used The *a priori* expectation for any particular face of the die is one-sixth, here one sixth of 16,000, or 2,666⅔ This is of the nature of a point binomial of the form $(p+q)^n$ with a standard deviation equal to $\sqrt{Npq}$

TABLE I

Distribution of Observed and Theoretical Populations with a Test of the Difference of Their Standard Deviations

Die Face	Observed Frequency	Expected Frequency	Difference
1	2726	2666⅔	59⅓
2	2653	2666⅔	14⅔
3	2671	2666⅔	4⅓
4	2763	2666⅔	96⅔
5	2650	2666⅔	17⅔
6	2537	2666⅔	130⅔

$\sigma = (1600 \cdot 1/6 \cdot 5/6)^{\frac{1}{2}} = 47.1$ $\qquad s = (\sum d^2/N)^{\frac{1}{2}} = 70.8$

$s - \sigma = 23.7 \pm 13.76$

Table I gives the observed and expected values of each face. The standard deviation of the differences was determined and compared with the standard deviation of the expected distribution and the probable error of this difference was found.

Small s is used here to denote a standard deviation of a sample, while σ represents the standard deviation of the theoretical or true population The formula for the standard deviation of a difference is

$$\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2 r_{12} \sigma_1 \sigma_2}$$

and in particular

$$\sigma_{s-d} = \sqrt{\sigma_s^2}$$

The second term drops out here because it is the standard deviation of the true standard error and this is equal to zero. The third term drops out for the same reason. Table I shows that the difference between the obtained and expected standard deviations is 23.7 ± 13.76. As this is less than twice its probable error, it can be concluded that the difference is not significant and that there is no significant bias in the dice.

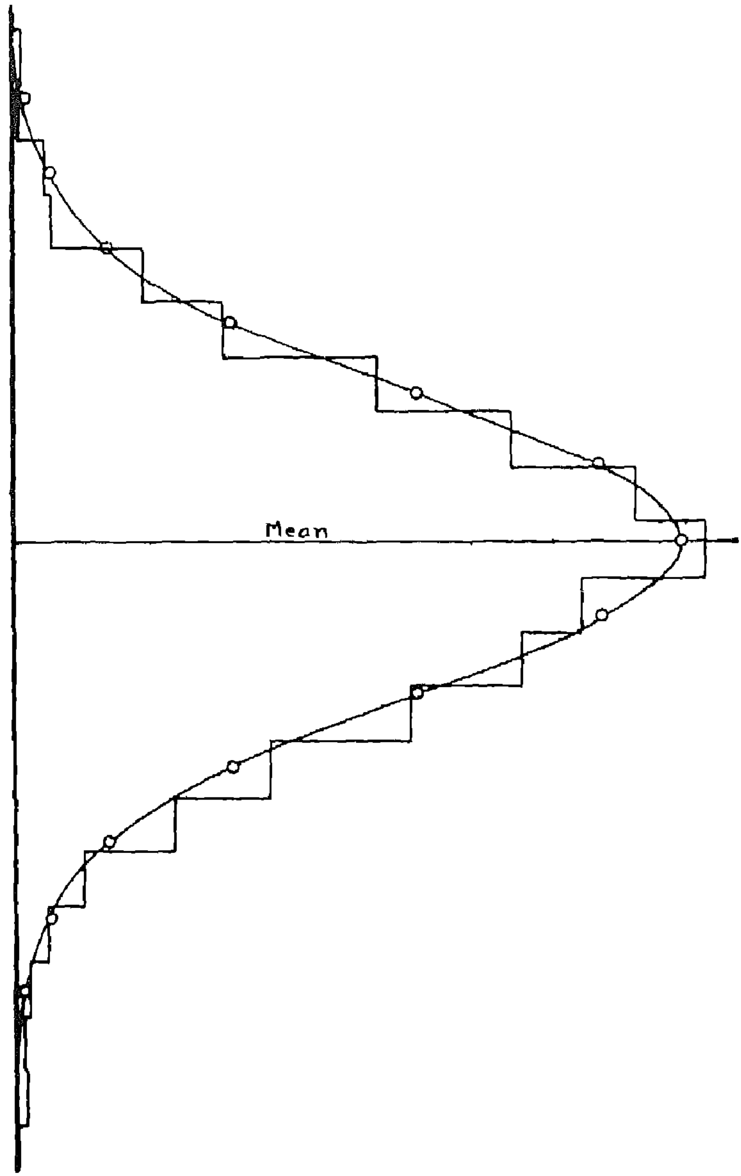
MEANS

Figure 1 shows the distribution of the 1600 observed means, a normal curve for $N = 1600$ is superimposed on the histogram. For this distribution

$$r_1 \left(= \sqrt{\beta_1} \right) = .0160 \pm .0413, \text{ indicating symmetry}$$

$$r_2 \left(\sqrt{\beta_2 - 3} \right) = -.1050 \pm .0826, \text{ indicating mesokurtosis}$$

whence we may conclude that the normal curve represents this


Mean

distribution adequately.

The curves of this and succeeding figures were drawn through points calculated at intervals of ½ σ, except that in the case of Figures 2 and 3, points beyond $\pm 2\sigma$ were calculated at intervals of 1σ.

The values of the observed means varied from 1.6 to 5.4, a range of 6.9129 standard deviations.

The basic information to be drawn from this study of the distribution of 1600 means of samples of ten is given in Table II. The table is interpreted as follows:

The mean of the sampled population (16,000) is 3.47306, while the theoretical mean of the infinite population is 3.500000. The standard deviation of the sampled population (16,000) is 1.6788, and of the theoretical population 1.7078. The standard error of the mean of the sampled population is .0133. In comparing the mean of the sampled population with the mean of the theoretical infinite population, the former is treated as an experimental value whose standard error can be estimated, while the latter, being a true value, has no error.

The standard deviation of the difference between the means M (theoretical population) and $\bar{X}$ (sampled population) is

$$\sigma_{(M-\bar{X})} = \sqrt{\sigma_M^2 + \sigma_{\bar{X}}^2 - 2\, r_{M\bar{X}}\, \sigma_M\, \sigma_{\bar{X}}}$$

$$\sqrt{\sigma_{\bar{X}}^2} = .0133$$

The first and third terms drop out because σ_M equals zero. The difference between the mean of the theoretical population and the sampled population is $.02694 \pm .00897$, from which it can be concluded that the mean tends to vary from the true mean.

$\bar{X}$ will hereafter refer to the mean of a sample of ten. The best estimate of the mean of a sample of ten that can be made for any sample chosen at random from the sampled population

TABLE II

Distribution of 1600 Means of Samples of 10

Description	Observed Value ($\bar{x}$)	Theoretical Value (M)
Mean of Sampled Pop.	3.47306	3.5000
σ of Sampled Pop.	1.6788	1.7078
σ_{mean} of Sampled Pop.	.0133	.0000
$\sigma_{(M-\bar{x})}$ of Sampled Pop.	.0133	.0000
$M-\bar{x}$ of Sampled Pop.	.0269 ± .00897	.0000
Mean of Means of Samples	3.47306	3.47306 or 3.5000
S. D. of Means of Samples	.5497	.5372 or .5401
S. E. of S. D. of Means of Samples	.0097	.0000 or .0000
$s_{\bar{x}} - \sigma_M$	.0125 ±.0065 or .0096 ± .0065	.0000 or .0000
γ_1 of Distri. of Means of Samples	.0160 ± .0413	.00 (normal theory)
γ_2 of Distri. of Means of Samples	-.1050 ± .0826	.00 (normal theory)

is 3.47306, and from the infinite population, 3.5000.

The standard deviation of the means of 1600 samples is .5467, while the estimated value for a sample picked at random from the sampled population is .5372 and from the theoretical infinite population .5401. These last two values are calculated by the formula

The best estimate of the standard deviation of a sample of ten picked at random from the sampled population is the σ of the sampled population, 1.6788, or of the theoretical infinite population, 1.7078, whence the values in the tables are obtained.

The standard error of the standard deviation of the means of samples is .0097 The standard error of the standard error σ_M of the mean of a sample of ten from the sampled and theoretical infinite populations is zero, as these are true values.

The difference between the standard deviation of the means and the standard error of such means of samples of ten from the sampled population or the theoretical infinite population is .0125 ±.0065. Thus there is no significant difference between the value of σ_M when calculated by the formula $\sigma_M = \frac{\sigma}{\sqrt{N}}$ and an actual distribution when samples as small as ten are used.

γ_1 indicates, as pointed out above, that the distribution is not skewed, while γ_2 shows the distribution to be slightly peaked but not significantly so.

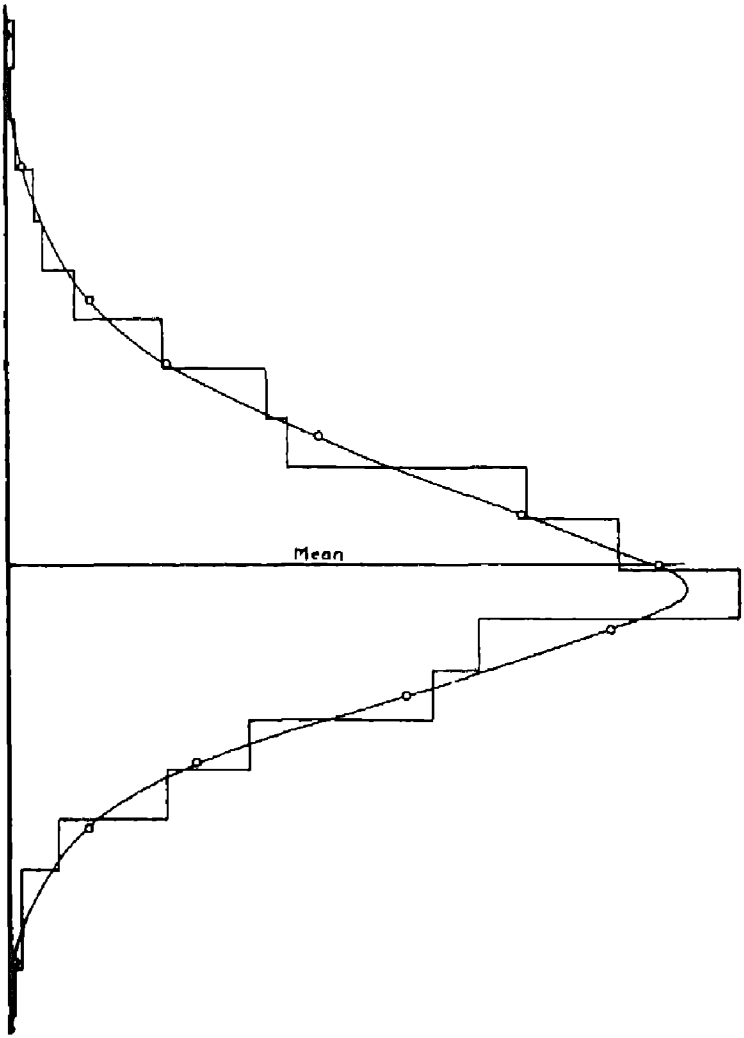
STANDARD DEVIATIONS

Figure 2 shows a histogram and a fitted Gram-Charlier Type A curve, of the distribution of 1600 standard deviations of samples of ten calculated by the formula

$$s = \sqrt{\frac{\sum x^2}{N}}$$

X being measured from the mean, $\bar{x}$

Figure 3 shows a similar histogram and curve fitted to the


Mean

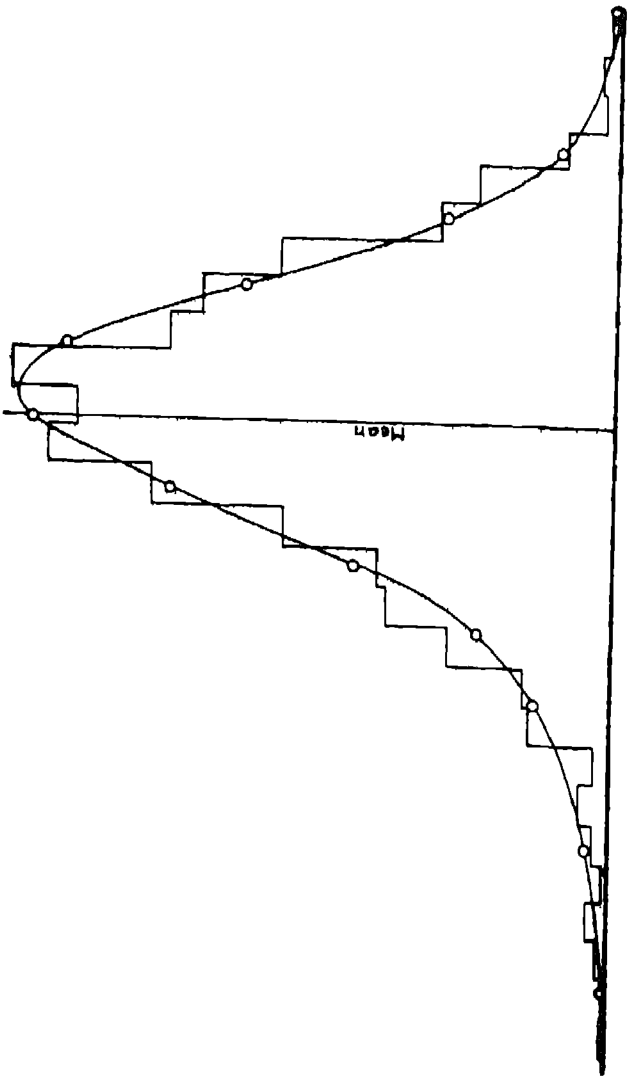


FIGURE 3

Distribution and fitted Gram-Charlier curve of 1600 standard deviations of samples of ten, calculated by the formula $s = \left(\frac{1}{N-1}\sum x^2\right)^{\frac{1}{2}}$

TABLE III

Distribution of 1600 Standard Deviations of Samples of Ten

Description	Observed Value		Theoretical Value	
	$s^2 = \frac{\Sigma x^2}{N}$	$s^2 = \frac{\Sigma x^2}{N-1}$	Sampled Population	Infinite Population
$\bar{x}$ of s's of sam.	1.5869	2.0403	1.6988	1.7078
S. D of s's of sam	.2665	.2538	.3799	.3818
S. D. of $\bar{x}$ of s's of samples	.0067	.0063	.0000	.0000
S. D. of s of s's of samples	.0047	.0045	.0000	.0000
$\sigma - \bar{x}_s$	.1119 ±.0045 or .1209 ±.0045	.3415 ±.0042 or .3325 ±.0042	.0000	.0000
$\sigma_\sigma - s_s$	.1134 ±.0032 or 1153 ±.0032	.1261 ±.0030 or .1280 ±.0030	.0000	.0000
γ_1 (skewness)	−.3568 ±.0413	−5026 ±.0413	.0000 (normal theory)	
γ_2 (kurtosis)	.5140 ±.0826	.6851 ±.0826	.0000 (normal theory)	
N	1600	1600		

same data when the standard deviations are calculated by the formula

$$s = \sqrt{\frac{\Sigma x^2}{N-1}}$$

A study of this latter formula is included here to test which is more appropriate when dealing with small samples from a rectangular population.

An interpretation of Table III is now in order. Column one is a description of the statistics involved. Column two is subdivided into two parts: First, when s equals $\sqrt{\frac{\Sigma x^2}{N}}$, and second when s equals $\sqrt{\frac{\Sigma x^2}{N-1}}$. Column three gives the theoretical values. There are two of these—one for the sampled population and one for the infinite population. In the case of the sampled population the values calculated for the standard deviation and the σ_σ become true values when a single sample is compared with them in exactly the same manner as if compared with similar values from the infinite population. The reason for this is that for a given sample the 16,000 constitutes the actual population from which the sample is drawn.

In the first line the means of the standard deviations of the samples are found to equal respectively, 1.5869 and 2.0403. The theoretical means for the sampled and infinite populations are respectively 1.6988 and 1.7078.

In the next line are the standard deviations of standard deviations of samples. These are calculated values, obtained by substituting in the formula

$$\sigma_s = \frac{\sigma}{\sqrt{2N}}$$

As the best estimate of the standard deviations of any particular sample chosen at random is the standard deviation of the sampled population, or the infinite population these values can be substituted in the above formula in obtaining the standard error of the standard deviation of such a sample of ten.

The standard error of the mean of standard deviations in

samples for both observed values is given in line three. Obviously in the case of the sampled and infinite populations these equal zero. It should be clearly understood by the reader that here N equals 1600, the number of standard deviations used in determining the mean standard deviation.

Line four gives the standard error of the standard deviation of standard deviations of samples of ten.

Line five gives the difference between each of the true standard deviations (sampled and infinite) and the two observed mean standard deviations The standard deviations of the sampled population and of the infinite population are each greater than the mean standard deviation of the observed population when calculated by the formula $s=\sqrt{\frac{\Sigma x^2}{N}}$. In the first case the difference is $1119 \pm .0045$. This is approximately 25 times its probable error, so it must be considered a significant difference. The difference when compared with the theoretical infinite population is $.1209 \pm .0045$ This is even more significant When the theoretical values are compared with the mean standard deviation calculated by the formula $s=\sqrt{\frac{\Sigma x^2}{N-1}}$ the differences are found to be $.3415 \pm .0042$, and $.3325 \pm .0042$. The differences here are much greater than those found from the first formula.

Line six shows the difference between the standard errors of the standard deviations of the true populations and the calculated s_s of the samples. The difference between σ_σ and s_s $(.3799 - .2665)$, is $.1134 \pm .0032$ This difference is approximately 35 times its probable error. The difference between .3799 and .2538 is even greater. Still larger differences are found when s_s is calculated for the $s=\sqrt{\frac{\Sigma x^2}{N-1}}$ formula.

γ_1 in the case of both curves is negative and more than 8 times its probable error, definitely showing a negative skewness. γ_2 in the case of both curves is 6 times greater than its probable error, indicating definite leptokurtosis. The Gram-Charlier curves shown in Figures 2 and 3 were fitted to the first four

moments according to the equation

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[1 - \left(\frac{\mu_3}{6\sigma^3}\right)(3x - x^3) + \left(\frac{\mu_4}{24\sigma^4} - .125\right)(x^4 - 6x^2 + 3) \right]$$

where $$x = \frac{X - \bar{X}}{s}$$

If we compute values of s by the empirical formula $s = \sqrt{\frac{\Sigma x^2}{N - .25}}$, the mean value is 1.7039, which lies very close to the theoretical values 1.6988 and 1.7078, in fact almost exactly half-way between them.

CORRELATION COEFFICIENTS

The product-moment correlation coefficient varies between the limits plus one and minus one. Obviously, the distribution of correlation coefficients cannot be normal, although in the case where $r = 0$ their distribution should approximate a normal curve, as it can become symmetrical. Coefficients around any other point tend to be distributed asymmetrically

It was assumed that if a deck of cards be thoroughly shuffled there should be no correlation between successive deals. Using a deck of cards gives a sample of 52. A new pack was thoroughly shuffled. The cards were then dealt one at a time, the first card dealt being recorded as number one, the second card dealt as number two, the third card as number three, etc, That is, if the seven of hearts was turned first, the value one was recorded against its place in the table. After each deal the cards were picked up in the same order and shuffled three times by the fan method and then cut twice. Sixty such deals were made and recorded. Then rank correlations were calculated be-

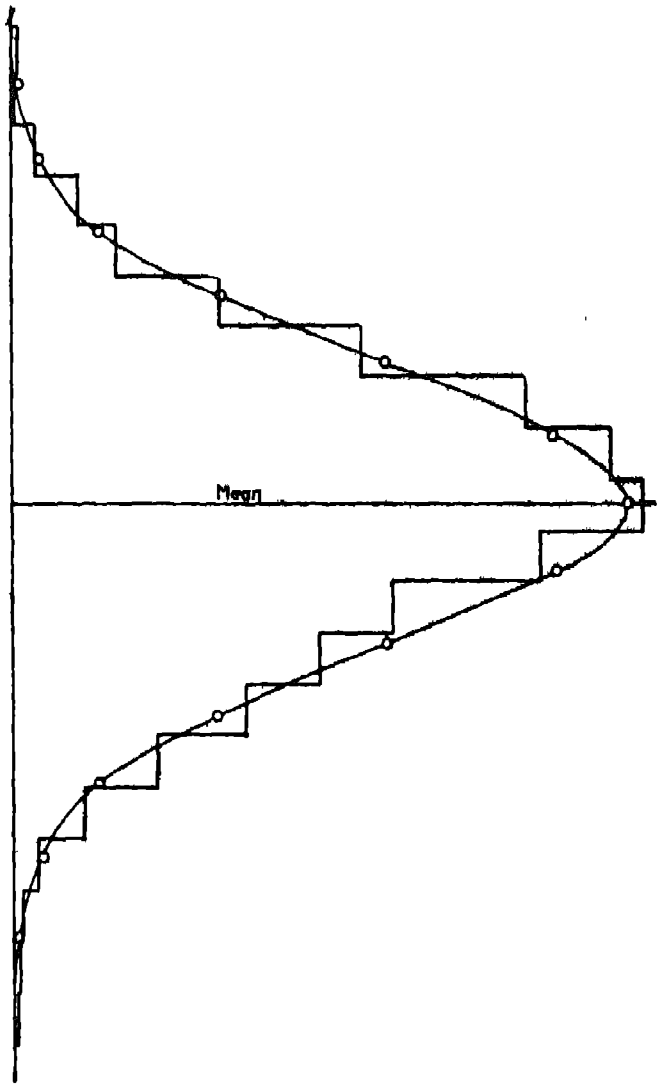


FIGURE 4
Distribution of 1770 correlation coefficients of samples of 52, with fitted normal curve.

tween each pair of deals, the total number of intercorrelations being $\frac{n(n-1)}{2}$, here 1770.

In this study, there could be no split ranks. Each card could receive one and only one rank on each deal. Thus, the rank correlation formula gave exactly the same values as would a Pearson product-moment coefficient.

Figure 4 shows a histogram with a fitted normal curve superimposed on it. γ_1 for this curve is .000015 ± .0392, indicating no skewness, and γ_2 is 2174 ± .0785, indicating a slight tendency to peakedness. Both of these facts are shown by the fit of the curve to the histogram.

The formula for the standard error of a correlation coefficient from a normal population is

$$\sigma_r = \frac{1-\rho^2}{\sqrt{N}}$$

ρ being the correlation in the population. Thus when $r = .0000$ and $N = 52$, $\sigma_r = .1387$.

The mean value of the 1770 coefficients is $r = -.0012$. The expected mean is zero. The difference between these two values is .0012 ± .0022. This shows that the mean correlation coefficient is not significantly different from the expected mean correlation.

The standard deviation of the observed distribution is .1359. This value differs from the expected value by .0028 ± .0091. The formula $\sigma_r = \frac{1-\rho^2}{\sqrt{N}}$ is therefore seen to give a sufficiently close approximation in this case.

CONCLUSIONS

1. The distribution of means of samples of ten drawn from a discontinuous rectangular population is normal. The formula $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ gives a reasonably close estimate of the standard error of such means.

2. The distribution of standard deviations of samples of

ten drawn from a discontinuous rectangular population is skewed and leptokurtic. The formula $\sigma_s = \frac{\sigma}{\sqrt{2N}}$ does not give a reasonably close estimate of the standard deviation of standard deviations of samples of ten, whether the latter are computed from the formula $s = \sqrt{\frac{\Sigma x^2}{N}}$ or $s = \sqrt{\frac{\Sigma x^2}{N-1}}$

3 Neither of the formulas, $s = \sqrt{\frac{\Sigma x^2}{N}}$ and $s = \sqrt{\frac{\Sigma x^2}{N-1}}$ for the standard deviation of a sample of ten gives a reasonably close estimate of the true standard deviation in a rectangular discontinuous population. The empirical formula $s = \sqrt{\frac{\Sigma x^2}{N-.25}}$ does appear to do so.

4. The distribution of correlation coefficients of samples of 52 from a rank population in which the expected correlation is zero, is symmetrical and very slightly leptokurtic. The formula $\sigma_r = \frac{1-\rho^2}{\sqrt{N}}$ represents adequately the standard deviation of such correlation coefficients.

Hilda F Dunlap

EDITORIAL

The Interdependence of Sampling and Frequency Distribution Theory

The object of the theory of sampling is to describe the phenomena exhibited by all the samples that can possibly arise from a parent population of known characteristics. In some cases the desired description can be obtained directly by employing elementary operations of combination theory, in others it is either expedient or necessary to use the indirect attack of the statistical theory of sampling. These two methods are quite different in application, and it is advisable to illustrate the respective peculiarities of the two methods.

Example 1. An auction bridge hand may be regarded as a single sample withdrawn from a parent population of 52 cards. The number of different hands that can be selected equals the number of combinations of 52 things taken 13 at a time, namely, $\binom{52}{13} = 635\ 013\ 559\ 600$. Of these

$$(1) \cdots\cdots \quad f(z) = \binom{39}{13-z}\binom{13}{z}$$

will contain exactly z cards of any specified suit. Therefore if in this expression we successively place z equal to 0, 1, 2, . . . 13 we shall obtain the frequency of all possible samples ranked according to the number of cards of the specified suit contained in each sample. The results are presented in the following table.

TABLE I

z	$f(z)$	$P_z = f(z)/N$
0	8 122 425 444	.01279
1	50 840 366 668	.08006
2	130 732 371 432	.20587
3	181 823 183 256	.28633
4	151 519 319 380	.23861
5	79 181 063 676	.12469
6	26 393 687 892	.04156
7	5 598 661 068	.00882
8	740 999 259	.00117
9	58 809 465	.00009
10	2 613 754	.00000
11	57 798	.00000
12	507	.00000
13	1	.00000
Total	635 013 559 600	.99999

In this illustration, combination theory has yielded a perfect solution. The frquencies are exact, and the sum of the frequencies between any two limits may likewise be obtained exactly by a simple addition.

Example 2. The bidding strength of hands in auction bridge is often approximated by counting each Jack, Queen, King and Ace as 1, 2, 3 and 4 points, respectively. The total count of a single hand may range, therefore from 0 to 37 inclusive. Required the frequency distribution of all possible hands when they

are classified according to count.

Unlike the preceding problem, we cannot obtain a simple expression for the general term, f_z, of the required distribution. But after rather involved computations the following solution may be obtained:

TABLE II

Count z	Frequency $f(z)$	Count z	Frequency $f(z)$
0	2 310 789 600	19	6 579 838 440
1	5 006 710 800	20	4 086 538 404
2	8 611 542 576	21	2 399 507 844
3	15 636 342 960	22	1 333 800 036
4	24 419 055 136	23	710 603 628
5	32 933 031 040	24	354 993 864
6	41 619 399 184	25	167 819 892
7	50 979 441 968	26	74 095 248
8	56 466 608 128	27	31 157 940
9	59 413 313 872	28	11 790 760
10	59 723 754 816	29	4 236 588
11	56 799 933 520	30	1 396 068
12	50 971 682 080	31	388 196
13	43 906 944 752	32	109 156
14	36 153 374 224	33	22 360
15	28 090 962 724	34	4 484
16	21 024 781 756	35	624
17	14 997 080 848	36	60
18	10 192 504 020	37	4
		Total	635 013 559 600

Example 3. If the mean and the standard deviation of the weights of a group of 200,000 men be 140 lbs. and 20 lbs., respectively, and if in addition it be known that the higher standard moments of this distribution be

$$\alpha_{3,x} = .5 \qquad \alpha_{5,x} = 4.43$$

$$\alpha_{4,x} = 3.17 \qquad \alpha_{6,x} = 17.97,$$

what is the chance that the mean weight of 1000 men chosen at random from the 200,000 will exceed 141 pounds?

It is clear that it would be physically impossible to solve this problem by employing a direct attack by combination theory, even though the weights of each of the 200,000 men were available. Moreover, it is likewise evident that in statistical problems corresponding to the illustrations of examples 1 and 2, the number of individuals in both the parent population and each sample is considerably larger than 52 and 13 respectively, and consequently the calculation of either a single frequency or the sum of any large group of consecutive frequencies by the direct method is quite out of the question.

Let us now consider the three examples above from the point of view of the indirect attack. The parent populations for the first two examples may be interpreted as

Variates	x	0 -	1
Frequencies	$f(x)$	39 -	13

and

Variates . .	x . .	0	1	2	3	4
Frequencies . .	$f(x)$. .	36	4	4	4	4

respectively.

For the first, the mean is at $x = 1/4$, and the moments about the mean of the parent population are obviously

$$\mu_{n:x} = \frac{13}{4^n}\left[3^n + 3(-1)^n\right]$$

For the second, the mean is at $x = 10/13$, and correspondingly the moments of this parent population are

$$\mu_{n:x} = \frac{1}{13^n}\left[(-10)^n + 3^n + 16^n + 29^n + 42^n\right]$$

If s and r denote the number of individuals in the parent population and each sample respectively, then the moments of the distribution of all samples that can arise from this parent population may be obtained from those of the parent population by means of the relations

$$(2)^1 \begin{cases} M_s = r \cdot M_x \\ \mu_{2:s} = \mu_{2:x}\, s(p_1 - p_2) \\ \mu_{3:s} = \mu_{3:x} \cdot s(p_1 - 3p_2 + 2p_3 \\ \mu_{4:s} = \mu_{4:x} \cdot s(p_1 - 7p_2 + 12p_3 - 6p_4) + 3\mu_{2:x}^2 \cdot s^2(p_2 - 2p_3 + p_4) \\ \mu_{5:s} = \mu_{5:x} \cdot s(p_1 - 15p_2 + 50p_3 - 60p_4 + 24p_5) \\ \qquad + 10\mu_{3:x}\,\mu_{2:x} \cdot s^2(p_2 - 4p_3 + 5p_4 - 2p_5) \\ \mu_{6:s} = \mu_{6:x} \cdot s(p_1 - 31p_2 + 100p_3 - 390p_4 + 360p_5 - 180p_6) \\ \qquad + 15\mu_{4:x}\,\mu_{2:x}\, s^2(p_2 - 8p_3 + 19p_4 - 18p_5 + 6p_6) \\ \qquad + 10\mu_{3:x}^2 \cdot s^2(p_2 - 6p_3 + 13p_4 - 12p_5 + 4p_6) \\ \qquad + 15\mu_{2:x}^3 \cdot s^3(p_3 - 3p_4 + 3p_5 - p_6) \end{cases}$$

where

$$\rho_i = \frac{r(r-1)(r-2)\cdots \text{to } i \text{ factors}}{s(s-1)(s-2)\cdots \text{to } i \text{ factors}}$$

Since the moments $\mu_{n\,x}$ for each of these three examples are now known, and according to the conditions of the problems the values of $(r,\ s)$ are (13, 52), (13, 52), and (1000, 200000) respectively, it follows that the moments of the desired distributions of samples are as follows:

Function	Example 1	Example 2	Example 3
$M_{\bar{x}}$	13/4	10	$M_{\bar{x}} = 140$ lbs.
$\mu_{2:\bar{x}}$	507/272	290/17	$\sigma_{\bar{x}} =$ .630874 lbs.
$\mu_{3:\bar{x}}$	6591/13600	288/17	$\alpha_{3:\bar{x}} =$ .0156927.
$\mu_{4:\bar{x}}$	53591421/5331200	17441114/29155	$\alpha_{4:\bar{x}} =$ 3.0001357
$\mu_{5:\bar{x}}$	9339447/1066240	2262240/833	$\alpha_{5:\bar{x}} =$ .1569051
$\mu_{6:\bar{x}}$	71781968037/801812480	2684384074/39151	$\alpha_{6:\bar{x}} =$ 15.026638

It will be observed that the indirect procedure has yielded the moments of the required distributions rather than their frequency functions, and the next step therefore is to obtain with the aid of these moments approximate expressions for the desired frequency functions. In this connection it should be borne in mind that we are not concerned with questions regarding the probable errors of the moments which we are employing, since the moments computed for the distributions of samples are necessarily exact, and their probable errors are therefore zero. For

[1]See Annals, Vol. I, page 104.

this reason arguments tending to limit the number of terms that may be employed in either a Gram-Charlier series, or in the denominator of Pearson's differential equation are not to the point so far as our illustrations are concerned. These remarks hold even for the third example, since if the moments of the parent population are as given, then the moments of the distribution of samples may be determined with any desired degree of accuracy.

Since it is evident that the solution of our problems now depends upon our obtaining approximate expressions for these distributions whose moments are known, we shall at this point develop a general method of representing discrete distributions which is essentially due to the researches of Charlier. Although the results that we shall obtain are practically those that have also been obtained by Gram, Edgeworth and others, the method that we shall employ is that used by Charlier in "Die Strenge Form des Bernoullischen Theorems."

Let $f(x)$ be the frequency function for a discrete distribution ranging from $x = \ell_1$ to $x = \ell_2$. If the ordinates be equidistant at intervals of h, the total frequency of the distribution is

$$(3)\quad N = f(\ell_1) + f(\ell_1 + h) + \cdots + f(x_0 - h) + f(x_0) + f(x_0 + h) + \cdots + f(\ell_2) = \sum_{x=\ell_1}^{\ell_2} f(x).$$

where our interest is focused on a typical ordinate at $x = x_0$. If we now set up the function

$$\sum_{x=\ell_1}^{\ell_2} f(x)\cdot e^{x\omega i} = f(x_0) + f(x_0+h)\cdot e^{(x_0+h)\omega i} + \cdots + f(\ell_2)e^{\ell_2\omega i} + f(x_0-h)e^{(x_0-h)\omega i} + \cdots + f(\ell_1)e^{\ell_1\omega i}$$

where $i=\sqrt{-1}$, and multiply each side by $e^{-x_0\omega i}$ so that

$$e^{-x_0\omega i}\sum_{x=l_1}^{l_2} f(x)e^{x\omega i} = f(x_0)+f(x_0+h)\,e^{h\omega i}+f(x_0+2h)\cdot e^{2h\omega i}+\cdots$$

$$+f(l_2)\cdot e^{(l_2-x_0)\omega i}+f(x_0-h)\,e^{-h\omega i}+f(x_0-2h)\,e^{-2h\omega i}+\ \ +f(l_1)e^{(l_1-x_0)\omega i}$$

we obtain by integrating both members with respect to ω between the limits $\omega=\frac{-\pi}{h}$ and $\omega=\frac{\pi}{h}$

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-x_0\omega i}\left\{\sum_{x=l_1}^{l_2} f(x)\,e^{x\omega i}\right\}d\omega=\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} f(x_0)\,d\omega,$$

since the integral of all other terms of the right hand member will vanish as follows:

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} f(x_0+mh)\cdot e^{mh\omega i}d\omega = f(x_0+mh)\cdot$$

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}\left[\cos mh\omega+i\sin mh\omega\right]d\omega=0$$

(m is an integer.)

It follows therefore that

(4) $$f(x_0)=\frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-x_0\omega i}\left\{\sum_{x=l_1}^{l_2} f(x)\,e^{x\omega i}\right\}d\omega$$

Moreover, since

$$e^{-a\omega i}+e^{-(a+h)\omega i}+\cdots+e^{-b\omega i}=\frac{e^{-(b+h)\omega i}-e^{-a\omega i}}{e^{-h\omega i}-1}$$

we see that the sum of all the consecutive frequencies from $x=a$ to $x=b$ may be expressed as the definite integral

$$(5)\quad \sum_{x=a}^{b} f(x)=\frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}\frac{e^{-(b+h)\omega i}-e^{-a\omega i}}{e^{-h\omega i}-1}\left\{\sum_{x=l_1}^{l_2} f(x)\cdot e^{x\omega i}\right\}d\omega$$

The changing of the order of integration is permitted since the limits are all finite.

Ordinarily frequency distributions are expressed as developments of the integral (4), and the sums of consecutive frequencies obtained by applying the Euler-Maclaurin Sum-Formula to these results. It seems at first sight that it might be well to place a little more emphasis upon the evaluation of (5), since this as it stands affords an exact expression for the sum of any group of consecutive frequencies. For the case of continuous variates we need only permit h to approach zero, replace the sign of summation by the sign of integration, etc., and after justifying the change in the order of integration for the resulting infinite limits obtain

$$(6)\quad \int_a^b f(x)dx=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-b\omega i}-e^{-a\omega i}}{-\omega i}\cdot\left\{\int_{l_1}^{l_2} f(x)\cdot e^{x\omega i}dx\right\}d\omega$$

We shall now attempt to evaluate the definite integral (4). Let us first observe that the quantity within the parenthesis is a function of ω, since the finite integration with respect to x and the subsequent replacing of x by the limits will cause this distribution variable to disappear.

For reasons which will develop later, let us write

$$\sum_{x=l_1}^{l_2} f(x)e^{x\omega i} = e^{b_1(\omega i)+b_2\frac{(\omega i)^2}{2!}} \sum_{x=l_1}^{l_2} f(x)\, e^{(x-b_1)\omega i - \frac{b_2(\omega i)^2}{2!}}$$

If in Leibnitz' formula

$$D^n u.v = u \cdot D^n v + \binom{n}{1} Du \cdot D^{n-1} v + \binom{n}{2} D^2 u \; D^{n-2} v + \cdots$$

we place $u = e^{\frac{bz^2}{2}}$ and $v = e^{az}$, and note that

$$D^{2n+1} e^{\frac{bz^2}{2}}\Big]_{z=0} = 0$$

$$D^{2n} e^{\frac{bz^2}{2}}\Big]_{z=0} = \frac{(2n)!}{2^n\, n!}\, b^n$$

then

$$(7)\quad D^n e^{az+\frac{bz^2}{2}}\Big]_{z=0} = a^n + \frac{n^{(2)}}{2 \cdot 1!} a^{n-2} b + \frac{n^{(4)}}{2^2 \cdot 2!} a^{n-4} b^2 + \frac{n^{(6)}}{2^3 \cdot 3!} a^{n-6} b^3 + \cdots$$

where $\quad n^{(i)} = n(n-1)(n-2) \; . \; . \; . \; . \; .$ to i factors.

Thus we may write

$$\sum_{x=\ell_1}^{\ell_2} f(x)\cdot e^{(x-b_1)\omega i - b_2\frac{(\omega i)^2}{2}} = N\left[c_0 + c_1(\omega i) + c_2\frac{(\omega i)^2}{2!} + c_3\frac{(\omega i)^3}{3!} + \cdots\right]$$

and employing the notation

$$\sum_{x=\ell_1}^{\ell_2} (x-b_1)^n f(x) = N\mu'_n$$

we obtain from (7)

$$(8)\quad \begin{cases} c_0 = 1 \\ c_1 = \mu'_1 \\ c_2 = \mu'_2 - b_2 \\ c_3 = \mu'_3 - 3b_2\mu'_1 \\ c_4 = \mu'_4 - 6b_2\mu'_2 + 3b_2^2 \\ \quad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ c_n = \mu'_n - \frac{n^{(2)}}{2\cdot 1!} b_2\mu'_{n-2} + \frac{n^{(4)}}{2^2 2!} b_2^2\mu'_{n-4} - \frac{n^{(6)}}{2^3 3!} b_2^3\mu'_{n-6} + \cdots \end{cases}$$

Formula (4) may therefore be written, dropping the subscript on x_0

$$(9)\qquad f(x) = N\frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-(x-b_1)\omega i - b_2\frac{(\omega i)^2}{2}}$$

$$\left[1 + c_1(\omega i) + c_2\frac{(\omega i)^2}{2} + \cdots\right] d\omega$$

Placing

$$(10)\qquad \Theta(x) = \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-(x-b_1)\omega i - \frac{b_2\omega^2}{2}}\, d\omega$$

it follows that the n th derivative with respect to x is

$$(11)\qquad \Theta^{(n)}(x) = \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} (-\omega i)^n \cdot e^{-(x-b_1)\omega i - \frac{b_2\omega^2}{2}}\, d\omega\,;$$

so finally

$$(12)\qquad f(x) = N\cdot h\left[\Theta(x) - \frac{c_1}{1!}\Theta^{(1)}(x) + \frac{c_2}{2!}\Theta^{(2)}(x) - \frac{c_3}{3!}\Theta^{(3)}(x) + \cdots\right]$$

Let us now investigate the function $\Theta(x)$.

$$\Theta(x) = \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-b_2\omega^2/2}\cdot\Big[\cos(x-b_1)\omega$$

$$- i\sin(x-b_1)\omega\Big]\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-b_2\omega^2/2}\cos(x-b_1)\,\omega\, d\omega$$

[since $e^{-b_2\omega^2/2}\sin(x-b_1)\omega$ is an odd function of ω]

$$=\frac{1}{\pi}\int_0^\infty e^{-b_2\omega^2/2}\cos(x-b_1)\omega\, d\omega$$

$$-\frac{1}{\pi}\int_{\frac{\pi}{h}}^\infty e^{-b_2\omega^2/2}\cos(x-b_1)\omega\, d\omega$$

$$=\frac{1}{\sqrt{2\pi b_2}}e^{-\frac{(x-b_1)^2}{2b_2}}-\frac{1}{\pi}\int_{\frac{\pi}{h}}^\infty e^{-b_2\omega^2/2}\cos(x-b_1)\omega\, d\omega$$

$$=\phi(x)-R_0\,.$$

$$\left[\int_0^\infty e^{-a^2x^2}\cos mx\,dx=\frac{\sqrt{\pi}}{a}\,e^{-m^2/4a^2}\right]$$

Likewise we may write

$$\theta^{(n)}(x)=\phi^{(n)}(x)-R_n\,,\qquad R_n<\frac{1}{\pi}\int_{\frac{\pi}{h}}^\infty \omega^n e^{-b_2\omega^2/2}\,d\omega$$

By successive integration by parts it can be shown that

$$\int x^n e^{-\frac{x^2}{2}}dx=-e^{-\frac{x^2}{2}}\Big\{x^{n-1}+(n-1)x^{n-3}$$

$$(13)\quad +(n-1)(n-3)x^{n-5}+\cdots+(n-1)(n-3)\cdots(n-2i+3)x^{n-2i+1}\Big\}+R_i\,,$$

$$R_i=(n-1)(n-3)\cdots(n-2i+1)\int x^{n-2i}e^{-\frac{x^2}{2}}\,dx$$

so we have that

$$(14)\quad R_n < \frac{1}{b_2}\left(\frac{\pi}{h}\right)^{n-1} e^{\frac{-b_2\left(\frac{\pi}{h}\right)^2}{2}}\left[1+\frac{n-1}{b_2}\left(\frac{h}{\pi}\right)^2+\frac{(n-1)(n-3)}{b_2^2}\left(\frac{h}{\pi}\right)^4+\cdots\right]$$

So far we have said nothing concerning the values of the parameters b_1 and b_2. Referring to formula (8) it is seen that if the origin of x be taken at the mean of the distribution in question, and b_2 equal the second moment about the mean of this distribution, $c_1 = c_2 = 0$, and consequently if the values of R_n may be neglected, the equation of the distribution expressed in standard units becomes

$$(15)\quad f(x) = N\frac{h}{\sigma}\left\{\phi(t) - \frac{A_3}{3!}\phi^{(3)}(t) + \frac{A_4}{4!}\phi^{(4)}(t) - \frac{A_5}{5!}\phi^{(5)}(t) + \cdots\right\}$$

where $t = \frac{x-b_1}{\sqrt{b_2}} = \frac{x-M}{\sigma}$, $\phi(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$, and

$$(16)\quad \begin{cases} A_3 = \alpha_3 \\ A_4 = \alpha_4 - 3 \\ A_5 = \alpha_5 - 10\alpha_3 \\ A_6 = \alpha_6 - 15\alpha_4 + 30 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ A_n = \alpha_n - \frac{n^{(2)}}{2\cdot 1!}\alpha_{n-2} + \frac{n^{(4)}}{2^2\cdot 2!}\alpha_{n-4} - \frac{n^{(6)}}{2^3\cdot 3!}\alpha_{n-6} + \cdots \end{cases}$$

By employing the Euler-Maclaurin Sum-Formula we can write

$$(17)\quad \begin{aligned} & f(a)+f(a+h)+f(a+2h)+\cdots+f(b-h)+f(b) \\ & = N\left[\int \phi(t)\,dt - A_0'\phi(t) + A_1'\phi^{(1)}(t) - A_2'\phi^{(2)}(t) + A_3'\phi^{(3)}(t)\cdots\right]_{\frac{a-M}{\sigma}}^{\frac{b+h-M}{\sigma}} \end{aligned}$$

where

$$(18)\quad \left\{ \begin{aligned} A_0' &= \frac{h}{2\sigma} \\ A_1' &= \frac{h^2}{12\sigma^2} \\ A_2' &= \frac{\alpha_3}{6} \\ A_3' &= \frac{\alpha_4-3}{24} + \frac{h}{\sigma}\,\frac{\alpha_3}{12} - \frac{h^4}{720\sigma^4} \\ A_4' &= \frac{\alpha_5-10\alpha_3}{120} + \frac{h}{\sigma}\,\frac{\alpha_4-3}{48} + \frac{h^2}{\sigma^2}\cdot\frac{\alpha_3}{72} \\ A_5' &= \frac{\alpha_6-15\alpha_4+30}{720} + \frac{h}{\sigma}\,\frac{\alpha_5-10\alpha_3}{240} + \frac{h^2}{\sigma^2}\cdot\frac{\alpha_4-3}{288} + \frac{h^6}{30240\sigma^6} \end{aligned} \right.$$

In some cases it may be more convenient to employ a mean and a standard deviation of the generating function that differs somewhat from that of the distribution for which the representation is desired. In this event the coefficients of the first and second derivatives in (15) will not vanish. However, the extra effort

expended in increasing the number of significant terms may be more than offset by the fact that a rather arbitrary choice in the values of b_1 and b_2 may result in simpler values for

$$t = \frac{x-b_1}{\sqrt{b_2}}$$

which in turn may occasionally eliminate difficult interpolations when dealing with tabulations of the generating function and its derivatives.

Formulae (17) and (18) may be regarded as a sort of apology for the fact that the definite integral of formula (5) has never been developed. The need of a satisfactory expression for the sum of any number of consecutive variates is indeed acute.

By permitting h in the foregoing theory to approach zero, one can obtain corresponding formulae for the ordinates and areas of distributions of continuous variates. However, it should be noted that for this case the limits for the integrals in the vicinity of formula (4) are now

$$\lim_{h \to 0} \frac{\pi}{h} = \infty$$

and consequently the changing of the order of integration must be justified.

In conclusion we may state:

I. Answers to problems of statistical sampling are usually expressed as finite or infinitesimal integrals under a function whose moments *only* are known. If known, the function is generally of but little value.

II. It is necessary to approximate the desired integrals by employing frequency functions.

III. Present methods are unsatisfactory from the point of view that remainder or limit of error terms are not available. The χ^2 test, though helpful, does not meet the issue in question.

H. C. Carver.

NOTE ON THE DISTRIBUTION OF MEANS OF SAMPLES OF N DRAWN FROM A TYPE A POPULATION

By

Cecil C. Craig
National Research Fellow

Recently in this journal, Dr. George A. Baker has found "the distribution of the means of samples drawn at random from a population represented by a Gram-Charlier series."[1] It is the purpose of this note to call attention to the fact that by the use of the semi-invariant notation Dr. Baker's results may be reached in very many fewer steps.

Let the parent population be represented by

$$(1)\quad f(x) = \phi(x)\left[1 + \frac{a_3}{\sigma_x^3} H_3\left(\frac{x}{\sigma_x}\right) + \frac{a_4}{\sigma_x^4} H_4\left(\frac{x}{\sigma_x}\right) + \cdots + \frac{a_k}{\sigma_x^k} H_k\left(\frac{x}{\sigma_x}\right)\right]$$

in which

$$(2)\qquad \phi(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma_x^2}}$$

[1]Vol. 1, No. 3 (Aug., 1930), pp. 199-204.

the origin for x being chosen at the mean, and

$$(3) \qquad H_k(t)e^{-\frac{t^2}{2}} = D_t^k\left(e^{-\frac{t^2}{2}}\right).$$

We shall first find the distribution function of $z = x_1 + x_2 + \cdots + x_N$ in which x_i, $i = 1, 2, \cdots N$, has the frequency function $f(x)$. Let us assume the frequency function of z is given by

$$(4) \qquad F(z) = \phi(z)\left[1 + \frac{A_3}{\sigma_z^3} H_3\left(\frac{z}{\sigma_z}\right) + \frac{A_4}{\sigma_z^4} H_4\left(\frac{z}{\sigma_z}\right) + \cdots + \frac{A_l}{\sigma_z^l} H_l\left(\frac{z}{\sigma_z}\right)\right]$$

Then the semi-invariants of $f(x)$, $\lambda_1, \lambda_2, \cdots \lambda_k$ are defined by the formal identity in t:

$$(5) \qquad e^{\lambda_1 t + \frac{1}{2!}\lambda_2 t^2 + \frac{1}{3!}\lambda_3 t^3 + \cdots} = \int_{-\infty}^{\infty} dx\, f(x) e^{xt} \quad (\lambda_1 = 0 \text{ in this case})$$

and on integration, using (3), we get at once on the right:

$$e^{\lambda_2 \frac{t^2}{2}}\left[1 - a_3 t^3 + a_4 t^4 + \cdots (-1)^k a_k t^k\right]$$

Similarly for the semi-invariants $L_1, L_2, L_3 \cdots$ of $F(z)$ we have

$$(6) \qquad e^{L_1 t + \frac{1}{2!}L_2 t^2 + \frac{1}{3!}L_3 t^3 + \cdots} = e^{L_2 \frac{t^2}{2}}\left[1 - A_3 t^3 + A_4 t^4 - \cdots + (-1)^l A_l t^l + \cdots\right]$$

But because of the well-known fact that $L_r = N\lambda_r$ this gives

$$1 - A_3 t^3 + A_4 t^4 - \cdots - (-1)^l A_l t^l$$
$$= \left[1 - a_3 t^3 + a_4 t^4 - \cdots (-1)^k a_k t^k\right]^N$$

an identity in t. Thus

$$(7)\quad A_r = \sum \frac{N!}{\nu_3!\,\nu_4!\cdots\nu_k!\,(N-\nu_3-\nu_4\cdots\nu_k)!}\, a_3^{\nu_3} a_4^{\nu_4}\cdots a_k^{\nu_k}$$

the summation including all terms for which

$$3\nu_3 + 4\nu_4 + \cdots + k\nu_k = r$$

Remembering that $\sigma_z = \sqrt{L_2} = \sqrt{N}\,\sigma_x$, we have on substitution in (4) the expression for $F(z)$ since only a finite number of A_r's (depending on N) are different from zero.

To get the distribution of $z' = \dfrac{x_1 + x_2 + \cdots + x_N}{N}$ only involves the appropriate change of unit.

Cecil C. Craig

Stanford University.

Vol. II No. 2

THE ANNALS
of
MATHEMATICAL STATISTICS

SIX DOLLARS PER ANNUM

MAY - 1931

ON SYMMETRIC FUNCTIONS AND SYMMETRIC FUNCTIONS OF SYMMETRIC FUNCTIONS*

By

A. L. O'TOOLE

INTRODUCTION

The study of symmetric functions is quite an old one. From the time of Girard (1629) even up to the present day this subject has occupied the attention of many eminent mathematicians. The theory of the roots of algebraic equations in one or more variables has furnished the chief incentive for the development of the theory of symmetric functions. Ingenious methods for computing symmetric functions in terms of what are called *the elementary symmetric functions* have been developed by Hammond, Brioschi, Junker, Dresden and others. Extensive tables of symmetric functions in terms of the elementary symmetric functions may be found in the literature.

Symmetric functions play such a pre-eminent rôle in the mathematical theory of statistics and their computation by direct methods or by general formulas, even when assumptions restricting the groupings of the variates about the various means are made, is so excessively tedious that there has seemed to be need

*A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University of Michigan.

of development of the theory of symmetric functions in directions not suggested by the theory of equations. The ingenious methods referred to above are of little or no practical value in statistics; for they express a symmetric function in terms of the elementary symmetric functions whilst here it is necessary to express the symmetric function in terms of what are called the *power sums.* Likewise, and for the same reason, the tables mentioned are of no value to the student of statistics.

Moreover, in the theory of sampling one not only has to deal with symmetric functions of the given variates but with symmetric functions of symmetric functions of the given variates. This then leads to interesting as well as practical developments in the theory of symmetric functions.

In this investigation it is proposed to:

1. Develop symbolic methods which will enable one to express any given symmetric function in terms of the power sums, without knowing the expressions for the symmetric functions of lower weight, and which will also lend themselves readily to the construction of tables;

2. Develop symbolic devices in the more general case of a symmetric function of symmetric functions.

CHAPTER I

Direct Computation

1. Suppose there is given a set of n variates[1] $x_1, \cdot x_2, x_3, x_4, \ldots x_n$, no assumptions whatever being made as to their arrangement about the various means. Any rational, integral, algebraic function of these n variates which is unaltered by interchanges or permutations of the variates is called *a symmetric function*. With a few modifications, the usual notation for symmetric functions will be used in this investigation.

The *power sums* $s_1, s_2, s_3, \ldots s_t$:

Let

$$s_1 = \sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots\cdots + x_n ,$$

$$s_2 = \sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + \cdots\cdots + x_n^2 ,$$

$$s_3 = \sum_{i=1}^{n} x_i^3 = x_1^3 + x_2^3 + \cdots\cdots + x_n^3 ,$$

$$\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots$$

$$s_t = \sum_{i=1}^{n} x_i^t = x_1^t + x_2^t + \cdots\cdots + x_n^t .$$

Further, let $(a^\alpha b^\beta c^\gamma \ldots)$ represent any symmetric

[1]The variates may be either real or complex numbers.

function of the given variates. In other words, let $(a^{\alpha} b^{\beta} c^{\gamma} \cdots)$ equal the sum of all the terms such as

$$x_1^a x_2^a \cdots x_{\alpha}^a x_{\alpha+1}^b x_{\alpha+2}^b \cdots x_{\alpha+\beta}^b x_{\alpha+\beta+1}^c \cdots x_{\alpha+\beta+\gamma}^c \cdots$$

which can be formed from the n variates, where $a, b, c, \ldots$ and $\alpha, \beta, \gamma, \ldots$ are positive integers and $a > b > c > \cdots > 0$. e. g.

$$(3^2 21) = \sum_{\substack{i=1\\ j=1\\ k=1\\ m=1}}^{n} x_i^3 x_j^3 x_k^2 x_m\,, \quad i \neq j \neq k \neq m.$$

DEFINITIONS:

A *partition* of a positive integer t is any set of positive integers whose sum is t. The integers which constitute the partition are called the *parts* of the partition and are enclosed in parentheses ( ). It is desirable to arrange the parts in descending order of magnitude from left to right. Obviously then for any finite positive integer t each partition of t contains a finite number of parts. If there are r parts in the partition of t then the partition is called an *r-part partition of t* or simply an *r-partition of t*. E. G. (33), (321), (3111) are respectively 2-part, 3-part and 4-part partitions of 6. When repeated parts appear in the partition it is customary to write one of the repeated parts with an index corresponding to the number of times that part is repeated. Thus (33) is written (3^2) and (3111) is written (31^3). The number t is called the *weight* of the partition. For a discussion of the formulae for finding the number of partitions of an integer the reader is referred to Whitworth's

"Choice and Chance."[1]

It will now be clear that the notation introduced for the general symmetric function is a partition notation. The *weight* of a symmetric function is the degree in all the variates of any term in the summation. The *order* of a symmetric function is the highest degree in which each variate appears in the summation. For instance, in $\sum x_i^4 x_j^3 x_k^2 =$ (432) the weight is $4+3+2=9$ and the order is 4. It follows that in the partition notation of a symmetric function the weight is given by $a\alpha + b\beta + c\gamma + \ldots$ and the order by a. In the partition notation the power sums become simply (1), (2), (3), ..., (r) respectively.

For the purpose of mathematical statistics, moments rather than the power sums are the important thing. However, the transformation from power sums to moments is so simple that the results of this investigation in terms of power sums may be written in terms of the moments by putting

$$\begin{aligned} n\mu'_{1:x} &= s_1 , \\ n\mu'_{2:x} &= s_2 , \\ n\mu'_{3:x} &= s_3 , \\ &\cdots\cdots \\ n\mu'_{r:x} &= s_r , \end{aligned}$$

where $\mu'_{1:x}$, $\mu'_{2:x}$, $\mu'_{3:x}$, . . . , $\mu'_{r:x}$ are the statistical *moments* of the n variates.

2. It is not difficult to express certain symmetric functions in terms of the power sums. Practically all texts in higher algebra devote a section or two to this problem. Most of those which develop general formulae do so by using the properties of the coefficients of an algebraic equation. However, many others have developed general formulae in symmetric functions without

[1] W. A. Whitworth, "Choice and Chance," G. E. Stechert and Co., N. Y., fifth edition, page 100.

making use of the algebraic equation in their derivations. The latter procedure will be followed here in order to emphasize the fact that the interest is not in the theory of equations but in a set of variates such as might appear for instance in a statistical problem. A few of the general formulae of symmetric functions will be developed now by direct computation in order to demonstrate a basic theorem of this work—a theorem which will be stated at the close of this chapter.

Multiplying s_2 and s_1 the result is

$$s_2 s_1 = (x_1^2+x_2^2+\cdots+x_n^2)(x_1+x_2+\cdots+x_n)$$

$$=(x_1^2x_2+x_1^2x_3+\cdots+x_{n-1}x_n^2)+(x_1^3+x_2^3+\cdots+x_n^3)$$

$$=\sum_{\substack{i=1\\j=1}}^{n} x_i^2 x_j + \sum_{i=1}^{n} x_i^3, \quad i \neq j$$

$(2)(1)=(21)+(3)$, hence

$(21) = (2)(1)-(3)$

Similarly, if $u \neq v$,

$$s_u s_v = (x_1^u + x_2^u + \cdots + x_n^u)(x_1^v + x_2^v + \cdots + x_n^v)$$

$$=(x_1^u x_2^v + x_1^u x_3^v + \cdots + x_{n-1}^u x_n^v) + (x_1^{u+v} + x_2^{u+v} + \cdots + x_n^{u+v})$$

$$= \sum_{\substack{i=1\\j=1}}^{n} x_i^u x_j^v + \sum_{i=1}^{n} x_i^{u+v} \quad , \ i \neq j \ ,$$

$= (uv) + (u+v)$, hence

$(uv) = (u)(v) - (u+v)$

However, if $u = v$ a modification is necessary. For then

$$(u)^2 = (x_1^u + x_2^u + \ldots + x_n^u)^2$$

$$= (x_1^{2u} + x_2^{2u} + \ldots + x_n^{2u}) + (x_1^u x_2^u + \ldots + x_{n-1}^u x_n^u)$$

$$= \sum_{i=1}^{n} x_i^{2u} + \sum_{\substack{i=1\\j=1}}^{n} x_i^u x_j^u \ , \quad i \neq j \ ,$$

$= (\overline{2u}) + 2(u^2)$ and thus

$2!(u^2) = (u)^2 - (\overline{2u})$ where the bar over $2u$ indicates ordinary algebraic multiplication of 2 and u , i e. $(\overline{2u}) = s_{2u}$.

If $u \neq v \neq w$, $u+v \neq w$, $u+w \neq v$, $v+w \neq u$, then

$$(u)(v)(w) = (x_1^u + x_2^u + \cdots + x_n^u)(x_1^v + x_2^v + \cdots + x_n^v)(x_1^w + x_2^w + \cdots + x_n^w)$$

$$= (x_1^u x_2^v x_3^w + \ldots) + (x_1^{u+v} x_2^w + \ldots) + (x_1^{u+v+w} + \ldots)$$

$$= \sum_{\substack{i=1\\j=1\\k=1}}^{n} x_i^u x_j^v x_k^w + \sum_{\substack{i=1\\j=1}}^{n} x_i^{u+v} x_j^w + \sum_{\substack{i=1\\j=1}}^{n} x_i^{v+w} x_j^u$$

$$+\sum_{\substack{i=1\\j=1}}^{n} x_i^{u+w} x_j^{v} + \sum_{i=1}^{n} x_i^{u+v+w}, \quad i \neq j \neq k$$

$$= (uvw) + (u+v, w) + (v+w, u) + (u+w, v) + (u+v+w)$$

the commas being used to separate the parts of the partitions. Now applying the result obtained for (uv) to the second, third and fourth terms on the right of this last expression, it becomes, since

$$(u+v, w) = (u+v)(w) - (u+v+w),$$

$$(v+w, u) = (v+w)(u) - (u+v+w),$$

$$(u+w, v) = (u+w)(v) - (u+v+w),$$

$$(u)(v)(w) = (uvw) + (u+v)(w) + (v+w)(u) + (u+w)(v) \cdot 2(u+v+w).$$

Finally

$$(uvw) = (u)(v)(w) - (u+v)(w) - (v+w)(u) - (u+w)(v) + 2(u+v+w)$$

$$= s_u s_v s_w - s_{u+v} s_w - s_{v+w} s_u - s_{u+w} s_v + 2s_{u+v+w}$$

If $u = v = w$, then a modification is again necessary, and repeating the multiplication with $u = v = w$ it is found that

$$3!(u^3) = (u)^3 - 3(\overline{2u})(u) + 2(\overline{3u})$$
$$= s_u^3 - 3s_{2u} s_u + 2s_{3u}.$$

In like manner, if $u \neq v \neq w \neq z$, $u+v \neq w$, etc., $u+v+w \neq z$; etc., then

$$(uvwz) = (u)(v)(w)(z) - (u)(v)(w+z) - (u)(w)(v+z)$$
$$-(u)(z)(v+w) - (v)(w)(u+z) - (v)(z)(u+w)$$
$$-(w)(z)(u+v) + 2(u)(v+w+z) + 2(v)(u+w+z)$$
$$+2(w)(u+v+z) + 2(z)(u+v+w) + (u+v)(w+z)$$
$$+(u+v)(w+z) + (u+w)(v+z) + (u+z)(v+w) - 6(u+v+w+z).$$

If $u = v = w = z$, then

$$4!(u^4) = (u)^4 - 6(u)^2(\overline{2u}) + 8(u)(\overline{3u}) + 3(\overline{2u})^2 - 6(\overline{4u})$$
$$= s_u^4 - 6s_u^2 s_{2u} + 8s_u s_{3u} + 3s_{2u}^2 - 6s_{4u}$$

Similar modifications are necessary when some but not all of the parts of the partition are equal. For example,

$$(u)^2(v) = (x_1^u + x_2^u + \cdots + x_n^u)^2 (x_1^v + x_2^v + \cdots + x_n^v)$$
$$= \sum_{i=1}^n x_i^{2u+v} + \sum_{\substack{i=1\\j=1}}^n x_i^{2u} x_j^v + \sum_{\substack{i=1\\j=1}}^n x_i^{u+v} x_j^u + 2\sum_{\substack{i=1\\j=1\\k=1}}^n x_i^u x_j^u x_k^v,$$
$$i \neq j \neq k,$$
$$= (\overline{2u}+v) + (\overline{2u}, v) + 2(u+v, u) + 2(u^2v)$$
$$= (\overline{2u})(v) + 2(u)(u+v) - 4(\overline{2u}+v) + 2(u^2v)$$

hence

$$2!(u^2v) = (u)^2(v) - (\overline{2u})(v) - 2(u)(u+v) + 2(\overline{2u}+v)$$

$$= s_u^2 s_v - s_{2u} s_v - 2s_u s_{u+v} + 2 s_{2u+v}$$

3. Proceeding after the above fashion, any symmetric function whatever can be expressed in terms of the power sums. However, the process becomes increasingly cumbersome and the general formula is of no practical value for the purpose of computation. Moreover, it is necessary to use a continuous process, that is, to work from the simpler symmetric functions of small weight to the more complex symmetric functions of greater weight.

A special case may be worth mentioning to illustrate still better the carrying out of the direct process in the general case.

$$(u)^t = (x_1^u + x_2^u + \cdots + x_n^u)^t$$

Applying the multinomial theorem and assuming that the law holds for $t-1$ and that the symmetric functions of weight less than t are known and transposing all the terms of the right member except the term involving (u^t), it is found that

$$t!(u^t) = \sum(-1)^{v+t} \frac{t!\,(u)^{a_1}(\overline{2u})^{a_2}(\overline{3u})^{a_3}\cdots(\overline{tu})^{a_t}}{1^{a_1}2^{a_2}3^{a_3}\cdots t^{a_t}\cdot a_1!\,a_2!\,a_3!\cdots a_t!}$$

where $a_1, a_2, a_3, \ldots, a_t$ are either positive integers or zeros such that $a_1 + a_2 + a_3 + \ldots + a_t = v$ and $a_1 + 2a_2 + 3a_3 + \cdots + ta_t = t$.

In particular, if $u = 1$, then

$$t!(1^t) = \sum (-1)^{v+t} \frac{t!(1)^{a_1}(2)^{a_2}(3)^{a_3}\cdots(t)^{a_t}}{1^{a_1}2^{a_2}3^{a_3}\cdots t^{a_t}\cdot a_1!\,a_2!\,a_3!\cdots a_t!}$$

This last result may be expressed very conveniently in determinant form. Starting with the results obtained in article 2, it is seen that

$$1!(1) = s_1 ,$$

$$2!(1^2) = \begin{vmatrix} s_1 & 1 \\ s_2 & s_1 \end{vmatrix}$$

$$3!(1^3) = \begin{vmatrix} s_1 & 1 & 0 \\ s_2 & s_1 & 2 \\ s_3 & s_2 & s_1 \end{vmatrix} ,$$

$$4!(1^4) = \begin{vmatrix} s_1 & 1 & 0 & 0 \\ s_2 & s_1 & 2 & 0 \\ s_3 & s_2 & s_1 & 3 \\ s_4 & s_3 & s_2 & s_1 \end{vmatrix} ,$$

. . . .

. . . .

$$t!(1^t) = \begin{vmatrix} s_1 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ s_2 & s_1 & 2 & 0 & \cdot & \cdot & \cdot & 0 \\ s_3 & s_2 & s_1 & 3 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ s_{t-2} & \cdot & \cdot & s_3 & s_2 & s_1 & t-2 & 0 \\ s_{t-1} & s_{t-2} & \cdot & \cdot & s_3 & s_2 & s_1 & t-1 \\ s_t & s_{t-1} & s_{t-2} & \cdot & \cdot & s_3 & s_2 & s_1 \end{vmatrix}$$

To establish this general law it is sufficient to note that the development of this determinant gives as a general term

$$(-1)^{v+t} \frac{t!\, s_1^{a_1}\, s_2^{a_2}\, s_3^{a_3} \cdots s_t^{a_t}}{1^{a_1}\, 2^{a_2}\, 3^{a_3} \cdots t^{a_t} \cdot a_1!\, a_2!\, a_3! \cdots a_t!}$$

where $a_1, a_2, a_3, \ldots, a_t$ are positive integers or zeros which satisfy the conditions $a_1 + a_2 + a_3 + \cdots + a_t = v$ and $a_1 + 2a_2 + 3a_3 + 4a_4 + \cdots ta_t = t$.

Hence the determinant is equal to

$$\sum (-1)^{v+t} \frac{t!\, s_1^{a_1}\, s_2^{a_2} \cdots s_t^{a_t}}{1^{a_1} 2^{a_2} \cdots t^{a_t} \cdot a_1!\, a_2! \cdots a_t!}$$

where, as before, the summation is over all the different terms it is possible to obtain by assigning $a_1, a_2, \ldots, a_t$ all positive integral values or zeros which satisfy the conditions

$$a_1 + a_2 + \cdots + a_t = v,$$
$$a_1 + 2a_2 + \cdots + ta_t = t.$$

4. This chapter will be concluded here with the statement of a very important theorem which may now be written and which will serve as a basis for the developments in the chapters to follow.

Basic Theorem:

Any symmetric function (defined in article 1) may be expressed as a rational, integral, algebraic function of the power sums.

Further, each term in the expression for the symmetric function in terms of the power sums is of the same weight as the symmetric function itself. Hence a term which does not arise from a partition of the weight of the symmetric function cannot appear in the expression in terms of the power sums.

CHAPTER II

A Differential Operator Method of Computing Symmetric Functions in Terms of the Power Sums

5. Consider a symmetric function $(a^{\alpha} b^{\beta} c^{\gamma} \ldots)$ of weight w of the variates $x_1, x_2, \ldots, x_n$. By the theorem demonstrated in chapter I and stated at the close thereof it is possible to write

$$(a^{\alpha} b^{\beta} c^{\gamma} \ldots) = f(s_1, s_2, \cdots, s_w)$$

where f stands for a rational, integral, algebraic function of the power sums $s_1, s_2, \ldots, s_w$, and where each term in f is of total weight w, i. e. *isobaric.*

In the preceding chapter the direct method of computing a symmetric function in terms of the power sums has been illustrated. But that method has two major disadvantages. In the first place, it is necessary to know the expressions in terms of the power sums of the symmetric functions of lower weight; and in the second place, it becomes altogether impractical for anything but the simplest cases. It is proposed to develop a method which will have neither of these disadvantages—in other words, to develop a method which will express any given symmetric function directly in terms of the power sums without knowing the expressions for the symmetric functions of lower weight, and which will not become too unwieldy. In addition, the method ought to lend itself readily to the construction of tables of symmetric functions in terms of the power sums.

The method developed here will be a differential operator method. It may be stated at the outset that many schemes for

determining differential operators which will do the work are possible. The writer has investigated a number of them. The operators developed here are given because they seem to satisfy best the demands just imposed on the method of computation. In fact, their simplicity and the directness with which they produce results indicate that they are the simplest differential operators that can be developed for the problem.

6. Suppose now that a new variate $x_{n+1} = k$ is introduced. What effect will it have on $(a^{\alpha} b^{\beta} c^{\gamma} \ldots)$ and on f? First consider $(a^{\alpha} b^{\beta} c^{\gamma} \ldots)$. Since all the variates enter the symmetric function in exactly the same way, new terms involving k in all the ways in which the other variates appear will be introduced. For example, if the original set of variates is x_1, x_2, x_3, x_4 and the original symmetric function $(32) = \sum x_i^3 x_j^2$, $i \neq j$, then this symmetric function is made up of the terms

$$\begin{array}{cccc} x_1^3 x_2^2 & x_2^3 x_1^2 & x_3^3 x_1^2 & x_4^3 x_1^2 \\ x_1^3 x_3^2 & x_2^3 x_3^2 & x_3^3 x_2^2 & x_4^3 x_2^2 \\ x_1^3 x_4^2 & x_2^3 x_4^2 & x_3^3 x_4^2 & x_4^3 x_3^2 \end{array}$$

Introducing a new variate $x_5 = k$, produces the new terms

$$\begin{array}{cccc} x_1^3 k^2 & x_2^3 k^2 & x_3^3 k^2 & x_4^3 k^2 \\ k^3 x_1^2 & k^3 x_2^2 & k^3 x_3^2 & k^3 x_4^2 \end{array}$$

or that is, produces $\sum k^3 x_i^2$ and $\sum x_i^3 k^2$. And since k is a constant with respect to the summation, these summations may be written $k^3 \sum x_i^2$ and $k^2 \sum x_i^3$, $i = 1, 2, 3, 4$.

Hence $\sum\limits_{\substack{i=1 \\ j=1}}^{4} x_i^3 x_j^2$ becomes $\sum\limits_{\substack{i=1 \\ j=1}}^{4} x_i^3 x_j^2 + k^3 \sum\limits_{i=1}^{4} x_i^2 + k^2 \sum\limits_{i=1}^{4} x_i^3$, $i \neq j$.

i. e., (32) becomes (32) $+ k^3(2) + k^2(3)$.

Similarly k must enter $(a^\alpha b^\beta c^\gamma \ . \ . \ . \ .)$ just as every other variate does. As a result new terms are produced and $(a^\alpha b^\beta c^\gamma \ . \ . \ . \ .)$ becomes $(a^\alpha b^\beta c^\gamma \ . \ . \ . \ .)$

$$+ k^a(a^{\alpha-1} b^\beta c^\gamma \cdots) + k^b(a^\alpha b^{\beta-1} c^\gamma \cdots)$$

$$+ k^c(a^\alpha b^\beta c^{\gamma-1} \cdots) + \cdots\cdots$$

Next find what happens to $f(s_1, s_2, \ . \ . \ . \ . \ , s_w)$ when the new variate $x_{n+1} = k$ is introduced. From the definition of the power sums it follows that

$$\begin{array}{lll}
s_1 & \text{becomes} & s_1 + k \ , \\
s_2 & \text{becomes} & s_2 + k^2 \ , \\
s_3 & \text{becomes} & s_3 + k^3 \ , \\
\cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots \\
s_t & \text{becomes} & s_t + k^t \ , \\
\cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots \\
s_w & \text{becomes} & s_w + k^w \ .
\end{array}$$

Hence $f(s_1, s_2, \ . \ . \ . \ ., s_w)$ becomes

$$f(s_1 + k, \ s_2 + k^2, \ \cdots, \ s_w + k^w).$$

Taylor's series for several variables is

$$\begin{aligned}
f(x+h, y+k, z+m, \cdots) = {} & f(x, y, z, \cdots) \\
& + (h\partial/\partial x + k\partial/\partial y + m\partial/\partial z + \cdots)f \\
& + (h\partial/\partial x + k\partial/\partial y + m\partial/\partial z + \cdots)^2 \frac{f}{2!} \\
& + (h\partial/\partial x + k\partial/\partial y + m\partial/\partial z + \cdots)^3 \frac{f}{3!} \\
& + \cdots\cdots\cdots
\end{aligned}$$

where the multiplication of operators is algebraic.

Applying Taylor's series to the function under consideration, the result is

$$
\begin{aligned}
f(s_1+k,\, s_2+k^2,\, \dots,\, s_w+k^w) &= f(s_1, s_2, \dots, s_w) \\
&+ (k\partial/\partial s_1 + k\partial/\partial s_2 + \dots + k^w\partial/\partial s_w)\, f \\
&+ (k\partial/\partial s_1 + k\partial/\partial s_2 + \dots + k^w\partial/\partial s_w)^2\, \frac{f}{2!} \\
&+ (k\partial/\partial s_1 + k\partial/\partial s_2 + \dots + k^w\partial/\partial s_w)^3\, \frac{f}{3!} \\
&\cdots\cdots\cdots\cdots \\
&+ (k\partial/\partial s_1 + k\partial/\partial s_2 + \dots + k^w\partial/\partial s_w)^w\, \frac{f}{w!},
\end{aligned}
$$

all other terms being identically zero.

Now let

$$
\begin{aligned}
&d_1 = \partial/\partial s_1\,,\quad d_2 = \partial/\partial s_2\,, \dots, \\
&d_v = \partial/\partial s_v\,, \dots,\quad v = 1, 2, 3, \dots, w.
\end{aligned}
$$

Then $d_1^2 = (\partial/\partial s_1)(\partial/\partial s_1) = \partial^2/\partial s_1^2$ and similarly $d_v^u = \partial^u/\partial s_v^u$

It is now possible to write

$$
\begin{aligned}
f(s_1+k,\, s_2+k^2,\, \dots,\, s_w+k^w) &= f \\
&+ (kd_1 + k^2d_2 + k^3d_3 + \dots + k^w d_w)\, f \\
&+ (kd_1 + k^2d_2 + k^3d_3 + \dots + k^w d_w)^2\, \frac{f}{2!} \\
&+ (kd_1 + k^2d_2 + k^3d_3 + \dots + k^w d_w)^3\, \frac{f}{3!} \\
&\vdots \\
&+ (kd_1 + k^2d_2 + k^3d_3 + \dots + k^w d_w)^w\, \frac{f}{w!}
\end{aligned}
$$

Multiplying out and collecting coefficients of powers of k, this becomes

$$f(s_1+k,\ s_2+k^2,\cdots,s_w+k^w)=(1+kD_1+k^2D_2+k^3D_3+\cdots+k^wD_w)f,$$

all other terms vanishing, where

$$(1)\quad\left[\begin{array}{l}
D_1=d_1,\\
2!D_2=d_1^2+2d_2,\\
3!D_3=d_1^3+6d_1d_2+6d_3,\\
4!D_4=d_1^4+12d_1^2d_2+24d_1d_3+12d_2^2+24d_4,\\
5!D_5=d_1^5+20d_1^3d_2+60d_1^2d_3+60d_1d_2^2+120d_1d_4+120d_2d_3+120d_5,\\
6!D_6=d_1^6+30d_1^4d_2+120d_1^3d_3+180d_1^2d_2^2+360d_1^2d_4\\
\qquad+720d_1d_2d_3+720d_1d_5+720d_2d_4+120d_2^3+360d_3^2+720d_6,\\
\text{etc.}
\end{array}\right.$$

Applying the multinomial theorem and then picking out the coefficient of k^t, the general term in this coefficient is found to be of the form

$$\frac{d_a^A\,d_b^B\,d_c^C\cdots}{A!\ B!\ C!\cdots}$$

where a, b, c, . . . and A, B, C, . . . are positive integers which satisfy the condition $aA+bB+cC+\cdots=t$.

Hence

$$t!\,D_t = \sum \frac{t!\,d_a^A\,d_b^B\,d_c^C}{A!\,B!\,C!} \qquad \text{where } aA + bB + cC + \cdots = t;$$

i. e. the sum of all the different terms which can be formed by assigning to $a, b, c, \ldots, A, B, C, \ldots$ all positive integral values which satisfy the condition $aA + bB + cC + \cdots = t$.

From the above relations it follows also that

$$(2)\quad \left\{\begin{aligned}
d_1 &= D_1, \\
2d_2 &= -(D_1^2 - 2D_2), \\
3d_3 &= (D_1^3 - 3D_1D_2 + 3D_3), \\
4d_4 &= -(D_1^4 - 4D_1^2D_2 + 2D_2^2 + 4D_1D_3 - 4D_4), \\
5d_5 &= (D_1^5 - 5D_1^3D_2 + 5D_1^2D_3 + 5D_1D_2^2 - 5D_2D_3 - 5D_1D_4 + 5D_5). \\
6d_6 &= -(D_1^6 - 6D_1^4D_2 + 6D_1^3D_3 - 6D_1^2D_4 + 9D_1^2D_2^2 - 12D_1D_2D_3 \\
&\qquad + 6D_1D_5 + 6D_2D_4 - 2D_2^3 + 3D_3^2 - 6D_6). \\
td_t &= (-1)^{t+1} \sum (-1)^{t+v} \frac{(v-1)!\,t\,D_a^A D_b^B D_c^C \cdots}{A!\,B!\,C!\cdots}
\end{aligned}\right.$$

where $a, b, c, \ldots; A, B, C, \ldots$ are positive integers and where the summation is over all the different terms which it is possible to obtain by assigning positive integral values to $a, b, c, \ldots; A, B, C, \ldots$ which satisfy the conditions $A + B + C + \cdots = v$, $aA + bB + cC + \cdots = t$.

7. Now since $(a^{\alpha} b^{\beta} c^{\gamma} \ldots) = f$, therefore replacing f by $(a^{\alpha} b^{\beta} c^{\gamma} \ldots)$ the effect of the introduction of

the new variate $x_{n+1} = k$ may be written

$$(1+kD_1+k^2D_2+\cdots+k^wD_w)(a^\alpha b^\beta c^\gamma\ldots) = (a^\alpha b^\beta c^\gamma\ldots)$$

$$+k^a(a^{\alpha-1}b^\beta c^\gamma\ldots)+k^b(a^\alpha b^{\beta-1}c\ldots)+k^c(a^\alpha b^\beta c^{\gamma-1}\ldots)$$

$$+\ldots\ldots$$

Equating coefficients of equal powers of k, it is obvious that

$$(3)\quad \begin{cases} D_a(a^\alpha b^\beta c^\gamma\ldots) = (a^{\alpha-1}b^\beta c^\gamma\ldots), \\ D_b(a^\alpha b^\beta c^\gamma\ldots) = (a^\alpha b^{\beta-1}c^\gamma\ldots), \\ D_c(a^\alpha b^\beta c^\gamma\ldots) = (a^\alpha b^\beta c^{\gamma-1}\ldots), \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ D_a^\alpha D_b^\beta D_c^\gamma\ldots(a^\alpha b^\beta c^\gamma\ldots) = 1, \quad \text{and also that} \\ D_r(a^\alpha b^\beta c^\gamma\ldots) = 0 \text{ if } r \text{ is not among } a, b, c; \\ \ldots\ldots \end{cases}$$

The relations between d and D given above enable one to express $(a^\alpha b^\beta c^\gamma\ldots)$ in terms of the power sums.

One particular case is worthy of mention. If 1 is not among $a, b, c, \ldots$ then $D_1(a^\alpha b^\beta c^\gamma\ldots) = 0$ and hence $d_1 f = 0$ and therefore also $d_1^2 f = d_1^3 f = \cdots = d_1^w f = 0$. In this case the operator relations may be written simply

$$(1') \quad \begin{cases} D_1 = 0, \\ D_2 = d_2, \\ D_3 = d_3, \end{cases}$$

$$\left[\begin{array}{l} 2!D_4 = d_2^2 + 2d_4, \\ \quad D_5 = d_2 d_3 + d_5, \\ 3!D_6 = d_2^3 + 6d_2 d_4 + 3d_3^2 + 6d_6, \\ \text{etc.} \end{array}\right.$$

and

$$(2') \quad \left[\begin{array}{l} \quad d_1 = 0, \\ \quad d_2 = D_2, \\ \quad d_3 = D_3, \\ 2!d_4 = 2D_4 - D_2^2, \\ \quad d_5 = D_5 - D_2 D_3, \\ 3!d_6 = 6D_6 - 3D_3^2 + 2D_2^3 - 6D_2 D_4, \\ \text{etc.} \end{array}\right.$$

Hence when 1 is not among $a, b, c, \ldots$ then s_1 cannot appear in the expression of $(a^\alpha b^\beta c^\gamma \ldots)$ in terms of the power sums, i. e. all the coefficients of terms involving s_1 vanish identically. But it must not be assumed that if $s_1 = 0$ then $d_1 f = 0$. Ordinarily this will not be true. It is necessary to find $\partial f / \partial s_1$ and in it set $s_1 = 0$. In statistics $s_1 = 0$ corresponds to the case where the variates are grouped about their arithmetic mean, i. e. so that $M_x = 0$.

8. The application of these operators d and D to the computation of a symmetric function in terms of the power sums will now be demonstrated. After that their use in the construction of tables will be considered.

Suppose it is desired to express (3^2) in terms of the power sums. The only terms which may appear are given by the partitions of 6. There are eleven partitions of 6. Hence let

$$(3^2) = a_1 s_1^6 + a_2 s_1^4 s_2 + a_3 s_1^3 s_3 + a_4 s_1^2 s_2^2 + a_5 s_1^2 s_4 + a_6 s_1 s_5$$

$$+ a_7 s_1 s_2 s_3 + a_8 s_2^3 + a_9 s_3^2 + a_{10} s_2 s_4 + a_{11} s_6 .$$

Since (3^2) does not contain 1 as a part, $D_1 = d_1 = 0$ and s_1 cannot appear on the right side of the above equation, i. e.

$$a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = a_7 = 0.$$

Now operate on the left side of the equation with D_2 and on the right with d_2.

$$D_2(3^2) = 0,$$

$$d_2 f = 3 a_8 s_2^2 + a_{10} s_4 ,$$

hence $0 = 3 a_8 s_2^2 + a_{10} s_4$ and therefore $a_8 = a_{10} = 0$. Operating on the left with D_3 and on the right with d_3 gives $a_9 = 1/2$ since $D_3(3^2) = (3)$ and $d_3 f = 2 a_9 s_3$, i. e. $s_3 = 2 a_9 s_3$. Operating on the left with $6 D_6$ and on the right with $d_2^3 + 6 d_2 d_4 + 3 d_3^2 + 6 d_6$ gives $0 = 6 a_9 + 6 a_{11}$ and thus $a_{11} = -1/2$. Hence

$$(3^2) = (s_3^2 - s_6)/2.$$

Similarly let

$$(31^3) = a_1 s_1^5 + a_2 s_1^3 s_2 + a_3 s_1^2 s_3 + a_4 s_1 s_2^2 + a_5 s_1 s_4 + a_6 s_2 s_3 + a_7 s_5 .$$

Operate on the right with d_1^2 and on the left with D_1^2.

This gives

$$s_3 = 20\, a_1 s_1^3 + 6\, a_2 s_1 s_2 + 2\, a_3 s_3$$

hence

$$a_1 = a_2 = 0, \quad a_3 = \frac{1}{2}$$

Operate on the right with $2d_2$, and on the left with $(D_1^2 - 2D_2)$. Then $-s_3 = 4\, a_4 s_1 s_2 + 2\, a_6 s_3$ and $a_4 = 0,\ a_6 = -\frac{1}{2}$

Operate on the right with $4d_4$ and on the left with $-(D_1^4 - 4D_1^2 D_2 + 2D_2^2 + 4\, D_1 D_3 - 4D_4)$. Then

$$-4\, s_1 = 4\, a_5 s_1\ ,\ a_5 = -1$$

Similarly, operating on the right with $5\, d_5$ and on the left with its equivalent in terms of D, the result is $5 = 5a_7\ ,\ a_7 = 1$. Hence

$$(31^2) = (s_1^2 s_3 - 2\, s_1 s_4 - s_2 s_3 + 2\, s_5)/2\ .$$

In the case of (3^2) the operations on the left were performed with D_1, D_2, D_3 and $6\,D_6$, and on the right with their equivalent expressions in terms of d_1, d_2, d_3, d_4, d_5, d_6, with $D_1 = d_1 = 0$. In the case of (31^2) the operations on the right were performed with d_1^2, $2d_2$, $4d_4$ and $5d_5$, and on the left with their equivalent expressions in terms of D_1, D_2, D_3, D_4, D_5. Obviously it is immaterial from a theoretical point of view which procedure is followed. For practical purposes it will usually be found that the procedure followed in the case of (31^2) is preferable.

9. The application of the operators to the construction of tables of symmetric functions in terms of the power sums will now be illustrated.

Weight 1:

1. $(1) = s_1$.

Weight 2:

1. $(2) = s_2$,

2. $(1^2) = a_1 s_1^2 + a_2 s_2$.

$$D_1(1^2) = d_1(a_1 s_1^2 + a_2 s_2),\ a_1 = 1/2.$$

$$2D_2(1^2) = (d_1^2 + 2d_2)(a_1 s_1^2 + a_2 s_2),\ a_2 = -a_1 = -1/2.$$

$$(1^2) = (s_1^2 - s_2)/2.$$

Weight 3:

For all the symmetric functions of weight 3 f will be of the form

$$f = a_1 s_1^3 + a_2 s_2 s_1 + a_3 s_3.$$

$$d_1 f = 3a_1 s_1^2 + a_2 s_2.$$

$$(d_1^3 + 6d_1 d_2 + 6d_3)f = 6(a_1 + a_2 + a_3).$$

1. $(3) = s_3$.

2. $(21) = s_2 s_1 - s_3$, since $D_1(21) = (2) = s_2$; therefore

   $a_1 = 0,\ a_2 = 1;\ 6D_3(21) = 0$ and hence $a_3 = -a_2 = -1$.

3. $(1^3) = (s_1^3 - 3s_2 s_1 + 2s_3)/6$ since $D_1(1^3) = (1^2)$

   and $(1^2) = (s_1^2 - s_2)/2$;

therefore $\quad a_1 = 1/6, \; a_2 = -1/2;$

$6D_3(1^3) = 0$ hence $a_3 = -a_1 - a_2 = 1/3.$

Weight 4:

For all the symmetric functions of weight 4 f will have the form

$$f = a_1 s_1^4 + a_2 s_1^2 s_2 + a_3 s_1 s_3 + a_4 s_2^2 + a_5 s_4.$$

$$d_1 f = 4a_1 s_1^3 + 2a_2 s_1 s_2 + a_3 s_3.$$

$$(d_1^2 + 2d_2) f = 2(6a_1 + a_2) s_1^2 + 2(a_2 + 2a_4) s_2.$$

$$(d_1^4 + 12 d_1^2 d_2 + 24 d_1 d_3 + 12 d_2^2 + 24 d_4) f = 24(a_1 + a_2 + a_3 + a_4 + a_5).$$

1. $(4) = s_4$

2. $(2^2) = (s_2^2 - s_4)/2$ since $D_1(2^2) = 0,$

$a_1 = a_2 = a_3 = 0; \quad 2D_2(2^2) = 2(2) = 2s_2,$

$a_4 = 1/2; \; 24D_4(2^2) = 0, \;\; a_5 = -1/2.$

3. $(31) = s_3 s_1 - s_4$ since $D_1(31) = (3) = s_3,$

$a_1 = a_2 = 0, \;\; a_3 = 1. \;\; 2D_2(31) = 0,$

$a_4 = 0; \; 24D_4(31) = 0, \;\; a_5 = -1.$

4. $(21^2) = (s_1^2 s_2 - 2s_1 s_3 - s_2^2 + 2s_4)/2$ since

$D_1(21^2) = (21) = s_2 s_1 - s_3, \; a_1 = 0, \; a_2 = 1/2, \; a_3 = -1;$

$$2D_2(21^2) = 2(1^2) = (s_1^2 - s_2),\ 2a_4 + a_2 = -1,$$

$$a_4 = -1/2\ ;\ 24D_4(21^2) = 0,\ a_5 = 1.$$

3. $(1^4) = (s_1^4 - 6s_1^2 s_2 + 8s_1 s_3 + 3s_2^2 - 6s_4)/24,$

since $\quad D_1(1^4) = (1^3) = (s_1^3 - 3s_2 s_1 + 2s_3)/6$

$$a_1 = 1/24,\ a_2 = -1/4,\ a_3 = 1/3\ ;\ 2D_2(1^4) = 0,$$

$$2a_4 = -a_2,\ a_4 = 1/8\ ;\ 24D_4(1^4) = 0,\ a_5 = -1/4.$$

Weight 5:

$$f = a_1 s_1^5 + a_2 s_1^3 s_2 + a_3 s_1^2 s_3 + a_4 s_1 s_2^2 + a_5 s_1 s_4 + a_6 s_2 s_3 + a_7 s_5.$$

$$d_1 f = 5a_1 s_1^4 + 3a_2 s_1^2 s_2 + 2a_3 s_1 s_3 + a_4 s_2^2 + a_5 s_4.$$

$$(d_1^2 + 2d_2) f = 2(10a_1 + a_2) s_1^3 + 2(3a_2 + 2a_4) s_1 s_2$$

$$+ 2(a_3 + a_6) s_3.$$

$$(d_1^5 + 20 d_1^3 d_2 + 60 d_1^2 d_3 + 60 d_1 d_2^2 + 120 d_1 d_4 + 120 d_2 d_3$$

$$+ 120 d_5) f = 120(a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7).$$

1. $(5) = s_5$

2. $(32) = s_3 s_2 - s_5$, since $D_1(32) = 0$,

$a_1=a_2=a_3=a_4=a_5=0$; $2D_2(32)=2(3)$,

$a_3+a_6=1$, $a_6=1$; $120d_5(32)=0$, $a_7=-a_6=-1$.

3. $(41)=s_4s_1-s_5$, since $D_1(41)=(4)$,

$a_1=a_2=a_3=a_4=0$, $a_5=1$; $2D_2(41)=0$, $a_6=0$;

$120d_5(41)=0$, $a_7=-1$.

4. $(2^21)=(s_2^2s_1-s_4s_1-2s_3s_2+2s_5)/2$, since

$D_1(2^21)=(2^2)$, $a_5=-1/2$, $a_1=a_2=a_3=0$,

$a_4=1/2$; $2D_2(2^21)=2(21)$, $a_6=-1$;

$120D_5(2^21)=0$, $a_7=1$.

5. $(31^2)=(s_3s_1^2-2s_4s_1-s_3s_2+2s_5)/2$, since

$D_1(31^2)=(31)$, $a_1=a_2=0$, $a_3=1/2$, $a_4=0$, $a_5=-1$;

$2D_2(31^2)=0$, $a_6=-a_3=-1/2$; $120D_5(31^2)=0$, $a_7=1$

6. $(21^3)=(s_2s_1^3-3s_3s_1^2-3s_2^2s_1+6s_4s_1+5s_3s_2-4s_5)/6$,

since $D_1(21^3)=(21^2)$, $a_1=0$, $a_2=1/6$,

$a_3=-1/2$, $a_4=-1/2$, $a_5=1$; $2D_2(21^3)=2(1^3)$,

$a_6=5/6$; $120D_5(21^3)=0$, $a_7=-2/3$.

7. $(1^5)=(s_1^5-10s_2s_1^3+20s_3s_1^2+15s_2^2s_1-30s_4s_1-20s_3s_2$

$+24s_5)/120$, since $D_1(1^5)=(1^4)$, $a_1=1/120$, $a_2=-1/12$, $a_3=1/6$,

$a_4=1/8$, $a_5=-1/4$; $2D_2(1^5)=0$, $a_6=-a_3=-1/6$; $120D_5(1^5)=0$, $a_7=1/5$.

Weight 6:

$$f = a_1 s_1^6 + a_2 s_2 s_1^4 + a_3 s_3 s_1^3 + a_4 s_2^2 s_1^2 + a_5 s_4 s_1^2$$

$$+ a_6 s_3 s_2 s_1 + a_7 s_5 s_1 + a_8 s_4 s_2 + a_9 s_2^3 + a_{10} s_3^2 + a_{11} s_6 \,.$$

$$d_1 f = 6a_1 s_1^5 + 4a_2 s_2 s_1^3 + 3a_3 s_3 s_1^2$$

$$+ 2a_4 s_2^2 s_1 + 2a_5 s_4 s_1 + a_6 s_3 s_2 + a_7 s_5 \,.$$

$$(d_1^2 + 2d_2) f = 2(15a_1 + a_2) s_1^4 + 2(6a_2 + 2a_4) s_2 s_1^2$$

$$+ 2(3a_3 + a_6) s_3 s_1 + 2(a_4 + 3a_9) s_2^2 + 2(a_5 + a_8) s_4$$

$$(d_1^3 + 6d_1 d_2 + 6d_3) f = 6(20a_1 + 4a_2 + a_3) s_1^3$$

$$+ 6(4a_2 + 4a_4 + a_7) s_2 s_1 + 6(a_3 + a_6 + 2a_{10}) s_3 \,.$$

$$(d_1^6 + 30 d_1^4 d_2 + 120 d_1^3 d_3 + 180 d_1^2 d_2^2 + 360 d_1^2 d_4 + 720 d_1 d_2 d_3$$

$$+ 720 d_1 d_5 + 720 d_2 d_4 + 120 d_2^3 + 360 d_3^2 + 720 d_6) f$$

$$= 720(a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10} + a_{11}).$$

1. $(6) = s_6$.

2. $(3^2) = (s_3^2 - s_6)/2$, since operating on this symmetric function with D_1 . 2 D_2 . 6 D_3 . 720 D_6 and comparing coefficients of the symmetric functions thus obtained with the result of the operations on f above gives

$$a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = a_7 = a_8 = a_9 = 0$$

$$a_{10} = 1/2 \;, \quad a_{11} = -1/2 \,.$$

3. $(2^3) = (s_2^3 - 3s_4 s_2 + 2s_6)/6$. For operating with D_1 and comparing coefficients of $D_1 (2^3) = 0$ with $d_1 f$ above gives $a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = a_7 = 0$.

Similarly, operating with $2\,D_2$ gives $a_8 = 1/6$, $a_9 = -1/2$. Operating with $6D_3$ gives $a_{10} = 0$. Operating with $720\,D_6$ gives $a_{11} = 1/3$.

4. $(42) = s_4 s_2 - s_6$.

5. $(51) = s_5 s_1 - s_6$.

6. $(321) = s_3 s_2 s_1 - s_5 s_1 - s_4 s_2 - s_3^2 + 2s_6$.

7. $(41^2) = (s_4 s_1^2 - 2s_5 s_1 - s_4 s_2 + 2s_6)/2$.

8. $(2^2 1^2) = (s_2^2 s_1^2 - s_4 s_1^2 - 4s_3 s_2 s_1 + 4s_5 s_1$

$\qquad + 5s_4 s_2 - s_2^3 + 2s_3^2 - 6s_6)/4$.

9. $(31^3) = (s_3 s_1^3 - 3s_4 s_1^2 - 3s_3 s_2 s_1$

$\qquad + 6s_5 s_1 + 3s_4 s_2 + 2s_3^2 - 6s_6)/6$.

10. $(21^4) = (s_2 s_1^4 - 4s_3 s_1^3 - 6s_2^2 s_1^2 + 12 s_4 s_1^2 + 20 s_3 s_2 s_1$

$\qquad - 16 s_5 s_1 - 18 s_4 s_2 + 3s_2^3 + 8s_3^2 + 16 s_6)/24$.

11. $(1^6) = (s_1^6 - 15s_2 s_1^4 + 40 s_3 s_1^3 + 45 s_2^2 s_1^2 - 90 s_4 s_1^2 - 120 s_3 s_2 s_1$

$\qquad + 144 s_5 s_1 + 90 s_4 s_2 - 15 s_2^3 + 40 s_3^2 - 120 s_6)/720$.

Note that only the four operator relations given above have been used in finding the expressions for all eleven symmetric functions of weight 6.

CHAPTER III

Symmetric Functions of Symmetric Functions
A Problem in Sampling

10. Consider again the n variates $x_1, x_2, \cdots, x_n$. Let $s_{1:x}, s_{2:x}, s_{3:x}, \cdots\cdots, s_{t:x}$ denote the power sums, the x subscript being introduced here to keep in the foreground the fact that the summation is with respect to x. Now raise each variate to the power m, where m is a positive integer. Thus a new set of variates is produced, viz. $x_1^m, x_2^m, \cdots, x_n^m$ Suppose now that samples, each containing r variates, $(r \leqq n)$, are drawn in all possible ways from these n new variates. Obviously there will be ${}_nC_r$ samples. Denote[1] them as follows:

$$z_1 = x_1^m + x_2^m + \cdots\cdots + x_r^m = \sum^{r:1} x^m,$$

$$z_2 = x_2^m + x_3^m + \cdots\cdots + x_{r+1}^m = \sum^{r:2} x^m,$$

$$z_3 = x_3^m + x_4^m + \cdots\cdots + x_{r+2}^m = \sum^{r:3} x^m,$$

$$\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots$$

$$z_{{}_nC_r} = x_{n-r+1}^m + \cdots \quad \cdot + x_n^m = \sum^{r:{}_nC_r} x^m,$$

[1]Notation suggested by Editorial, *Annals of Mathematical Statistics*, 1 (1930), page 100.

where $z_i = \sum^{r:i} x^m$ is the sum of the r variates appearing in the i'th sample.

Further, let

$$s_{1:z} = \sum_{i=1}^{{}_nC_r} z_i\,,$$

$$s_{2:z} = \sum_{i=1}^{{}_nC_r} z_i^2\,,$$

$$s_{3:z} = \sum_{i=1}^{{}_nC_r} z_i^3\,,$$

$$\cdots\cdots\cdots$$

$$\cdots\cdots\cdots$$

$$s_{t:z} = \sum_{i=1}^{{}_nC_r} z_i^t$$

represent the power sums with respect to z.

Now, since each z_i is a symmetric function of certain of the $x_1^m, x_2^m, \ldots, x_n^m$, any symmetric function of the z_i is a symmetric function of symmetric functions. The situation here is then considerably more complex than in the preceding chapters. The problem now is to express any symmetric function of the z_i in terms of the power sums with respect to x. It is not difficult to imagine how much more complicated and tedious the direct computation is here than in the problem already dealt with. But these symmetric functions, particularly the power sums with respect to z, play such an important rôle in the theory of sampling that it is now proposed to develop a differential operator method for expressing symmetric functions of the z_i in terms of the power sums with respect to x.

On account of the presence here of symmetric functions of both x and z it is necessary to modify the notation of the pre-

ceding chapters. Let $(a^{\alpha} b^{\beta} c^{\gamma} \ldots)_x$ be the general symmetric function with respect to x and $(a^{\alpha} b^{\beta} c^{\gamma} \ldots)_z$ the same general symmetric function with respect to z. Under this notation the power sums with respect to x may be written $(1)_x$, $(2)_x$, . . ., $(t)_x$, and the power sums with respect to z become $(1)_z$, $(2)_z$, . . ., $(t)_z$.

11. Case $m=1$.

Consider first of all the case of samples when $m=1$. In developing an operator method for expressing $(a^{\alpha} b^{\beta} c^{\gamma} \ldots)_z$ in terms of the power sums with respect to x it will not be necessary to deal with this general case. For the operators developed in chapter II will express $(a^{\alpha} b^{\beta} c^{\gamma} \ldots)_z$ in terms of the power sums with respect to z. Hence all that is required is an operator method for expressing the power sums with respect to z in terms of the power sums with respect to x.

That it is possible to express the power sums with respect to z in terms of the power sums with respect to x can be demonstrated by direct methods. Recall the theorem stated at the close of chapter I and note also that in any power sum with respect to z each term is a symmetric function (a power sum in fact) of certain of the x_1, x_2, . . ., x_n. Each x enters exactly the same as every other x and the power sum with respect to x is unaltered by interchanges or permutations of x_1, x_2, . . ., x_n. Hence the symmetric function with respect to z, is also a symmetric function with respect to x and therefore can be expressed as a rational, integral, algebraic function of the power sums with respect to x. Moreover, as before, each term in the rational, integral, algebraic function of the power sums with respect to x will be of total weight w if the symmetric function of the z_i is of weight w; that is, the symmetric function is of the same weight in x as it is in z. This last conclusion follows directly from the definition of the z_i.

Although the problem here is more complicated than that

in chapter II, nevertheless the approach to the problem in that case suggests a beginning here. Let

$$(w)_z = f(s_{1:x}, s_{2:x}, \ldots, s_{w:x}),$$

where f is a rational, integral, algebraic function of the power sums with respect to x. Since $(w)_z$ is of weight w, no power sum of weight greater than w can appear in f, i. e. no power sum higher than $s_{w:x}$.

Introducing a new variate $x_{n+1} = k$, as before, changes $f(s_{1:x}, s_{2:x}, \ldots, s_{w:x})$ into $f(s_{1:x} + k, s_{2:x} + k^2, \ldots, s_{w:x} + k^w)$. But it has already been shown that this new f may be written

$$f(s_{1:x}+k, s_{2:x}+k^2, \cdots, s_{w:x}+k^w) = (1+kD_1 + k^2D_2 + \cdots + k^wD_w)f$$

where, if $d_v = \partial/\partial s_{v:x}$, the relations between D and d are given by (1) and (2) of chapter II.

What is the effect of the new variate $x_{n+1} = k$ on $(w)_z$? If no further assumptions are made then obviously there will now be ${}_{n+1}C_r$ samples. The introduction of new samples complicates things and no operator relations are obtained. It would seem desirable to preserve the number of samples. This may be done by making suitable assumptions. Just as the new variate is arbitrarily introduced, so its behaviour in the sampling process may be arbitrarily determined in any way that will bring results. With this in mind, select any one of the original variates, say x_i. Let $qx_i = k = x_{n+1}$. Now assume that $k = qx_i$ is so related with x_i that in the sampling process every sample which contains x_i also contains qx_i, i. e. contains $(q+1)x_i$. In other words, in order to keep the number of samples the same, x_i and qx_i are always taken together in the samples.

Now each variate appears in $(1)_z$ exactly ${}_{n-1}C_{r-1}$

times. Hence $(q+1)x_i$ appears ${}_{n-1}C_{r-1}$ times in the new $(1)_z$. Therefore the new $(1)_z$ is equal to the original $(1)_z$ increased by $qx_i \cdot {}_{n-1}C_{r-1} = k \cdot {}_{n-1}C_{r-1}$. Similarly $(2)_z$ becomes $(2)_z + 2k(1)_{z'} + k^2 \cdot {}_{n-1}C_{r-1}$ where the prime above z indicates here, and in what follows, that $(t)_{z'}$ is obtaind from $(t)_z$ by replacing n and r by $n-1$ and $r-1$ respectively in the expression for $(t)_z$ in terms of the power sums with respect to x. For example, since $(1)_z = {}_{n-1}C_{r-1} \cdot s_{1:x}$, then

$$(1)_{z'} = {}_{n-2}C_{r-2} \cdot s_{1:x}.$$

Applying the multinomial theorem to the samples, the effect of the new variate may be written

$(1)_z$ becomes $(1)_z + k \cdot {}_{n-1}C_{r-1}$,

$(2)_z$ becomes $(2)_z + 2k(1)_{z'} + k^2 \cdot {}_{n-1}C_{r-1}$,

$(3)_z$ becomes $(3)_z + 3k(2)_{z'} + 3k^2(1)_{z'} + k^3 \cdot {}_{n-1}C_{r-1}$,

. . . .

$$\begin{aligned}(w)_z \text{ becomes } (w)_z &+ {}_wC_1 \cdot k(w-1)_{z'} + {}_wC_2 \cdot k^2(w-2)_{z'} \\ &+ \cdots \\ &+ {}_wC_v \cdot k^v(w-v)_{z'} + \cdots \\ &+ \cdots \\ &+ {}_wC_{w-1} \cdot k^{w-1}(1)_{z'} + k^w \cdot {}_{n-1}C_{r-1}.\end{aligned}$$

Now since $(w)_z = f$, therefore

$$
\begin{aligned}
(1+kD_1+k^2D_2+\cdots+k^wD_w)(w)_z = (w)_z &+ {}_wC_1\cdot k(w-1)_{z'} \\
&+ {}_wC_2\cdot k^2(w-2)_{z'} + {}_wC_3\cdot k^3(w-3)_{z'} \\
&+ \cdots \quad \cdots \\
&+ {}_wC_v\cdot k^v(w-v)_{z'} + \cdots \\
&+ \cdots \quad \cdots \\
&+ k^w\cdot {}_{n-1}C_{r-1}
\end{aligned}
$$

Equating coefficients of equal powers of k it follows that:

$$D_1(w)_z = {}_wC_1\cdot(w-1)_{z'},$$

$$D_2(w)_z = {}_wC_2\cdot(w-2)_{z'},$$

$$D_3(w)_z = {}_wC_3\cdot(w-3)_{z'},$$

$$\cdots\cdots\cdots$$

$$D_v(w)_z = {}_wC_v\cdot(w-v)_{z'},$$

$$\cdots\cdots\cdots$$

$$D_{w-1}(w)_z = {}_wC_{w-1}\cdot(1)_{z'},$$

$$D_w(w)_z = {}_{n-1}C_{r-1}.$$

$$D_u(w)_z = 0 \quad \text{if} \quad u>w.$$

12. Before proceeding to the application of these operators it ought to be remarked that other sets of differential operators

can be developed. For instance, it is possible to develop a complete set of differential operator relations by adding k to each of the given variates. But the operators thus obtained are very cumbersome in comparison with those developed above. The statement made with respect to the operators developed in chapter II may be repeated here. There is every reason to believe that the differential operators developed here are the simplest that can be obtained for the problem.

13. The use of the operators developed in this chapter will now be illustrated by computing a few power sums with respect to z in terms of the power sums with respect to x.

1. Let $(1)_z = a_1 s_{1:x}$. Then

$$D_1 (1)_z = d_1 a_1 s_{1:x} \quad \text{, that is}$$

$$_{n-1}C_{r-1} = a_1 . \qquad \text{Hence}$$

$$(1)_z = {}_{n-1}C_{r-1} \; s_{1:x} .$$

2. Let

$$(2)_z = f = a_1 s_{1:x}^2 + a_2 s_{2:x} .$$

$$D_1 (2)_z = d_1 f ,$$

$$2 (1)_{z'} = d_1 f ,$$

$$2 \cdot {}_{n-2}C_{r-2} \cdot s_{1:x} = 2 a_1 s_{1:x} , \quad a_1 = {}_{n-2}C_{r-2} .$$

$$2! \, D_2 (2)_z = (d_1^2 + 2 d_2) f ,$$

$$2 \; {}_{n-1}C_{r-1} = 2 (a_1 + a_2) , \quad a_2 = {}_{n-1}C_{r-1} - {}_{n-2}C_{r-2} ,$$

$$(2)_z = {}_{n-2}C_{r-2} \; s_{1:x}^2 + ({}_{n-1}C_{r-1} - {}_{n-2}C_{r-2}) \, s_{2:x} .$$

3 Let

$$(3)_z = f = a_1 s_{1:x}^3 + a_2 s_{1:x} s_{2:x} + a_3 s_{3:x}.$$

$$D_1 (3)_z = d_1 f,$$

$$3(2)_{z'} = d_1 f,$$

$$3 \cdot {}_{n-3}C_{r-3} \cdot s_{1:x}^2 + 3({}_{n-2}C_{r-2} - {}_{n-3}C_{r-3})\, s_{2:x}$$

$$= 3a_1 s_{1:x}^2 + a_2 s_{2:x}, \quad \text{hence} \quad a_1 = {}_{n-3}C_{r-3},$$

$$a_2 = 3({}_{n-2}C_{r-2} - {}_{n-3}C_{r-3}).$$

$$3!\, D_3 (3)_z = (d_1^3 + 6 d_1 d_2 + 6 d_3) f,$$

$$6 \cdot {}_{n-1}C_{r-1} = 6(a_1 + a_2 + a_3),$$

$$a_3 = {}_{n-1}C_{r-1} - 3\, {}_{n-2}C_{r-2} + 2 \cdot {}_{n-3}C_{r-3}.$$

$$(3)_z = {}_{n-3}C_{r-3}\, s_{1:x}^3 + 3({}_{n-2}C_{r-2} - {}_{n-3}C_{r-3})\, s_{1:x} s_{2:x}$$

$$+ ({}_{n-1}C_{r-1} - 3 \cdot {}_{n-2}C_{1-2} + 2 \cdot {}_{n-3}C_{r-3})\, s_{3:x}.$$

4. Let

$$(4)_z = f = a_1 s_{1:x}^4 + a_2 s_{1:x}^2 s_{2:x} +$$

$$a_3 s_{1:x} s_{3:x} + a_4 s_{2:x}^2 + a_5 s_{4:x}$$

$$D_1 (4)_z = d_1 f.$$

$$4(3)_{z'} = d_1 f,$$

$$4[{}_{n-4}C_{r-4}\, s_{1:x}^3 + 3({}_{n-3}C_{r-3} - {}_{n-4}C_{r-4})\, s_{1:x} s_{2:x}$$

$$+({}_{n-2}C_{r-2}-3\cdot{}_{n-3}C_{r-3}+2\cdot{}_{n-4}C_{r-4})s_{3:x}]$$

$$=4a_1 s_{1:x}+2a_2 s_{1:x}s_{2:x}+a_3 s_{3:x},$$

$$a_1={}_{n-4}C_{r-4},\quad a_2=6({}_{n-3}C_{r-3}-{}_{n-4}C_{r-4}),$$

$$a_3=4({}_{n-2}C_{r-2}-3\cdot{}_{n-3}C_{r-3}+2\cdot{}_{n-4}C_{r-4}).$$

$$2!D_2(4)=(d_1^2+2d_2)f,$$

$$12(2)_{\Sigma'}=(d_1^2+2d_2)f,$$

$$12\cdot{}_{n-3}C_{r-3}\cdot s_{1:x}^2+12({}_{n-2}C_{r-2}-{}_{n-3}C_{r-3})s_{2:x}$$

$$=2(6a_1+a_2)s_{1:x}^2+2(a_2+2a_4)s_{2:x},$$

$$a_4=3({}_{n-2}C_{r-2}-2\cdot{}_{n-3}C_{r-3}+{}_{n-4}C_{r-4}).$$

$$4!D_4(4)_{\Sigma}=(d_1^4+12d_1^2d_2+24d_1d_3+12d_2^2+24d_4)f,$$

$$24\cdot{}_{n-1}C_{r-1}=24(a_1+a_2+a_3+a_4+a_5),$$

$$a_5={}_{n-1}C_{r-1}-7\cdot{}_{n-2}C_{r-2}+12\cdot{}_{n-3}C_{r-3}-6\cdot{}_{n-4}C_{r-4}.$$

$$(4)_{\Sigma}={}_{n-4}C_{r-4}\cdot s_{1:x}^4+6({}_{n-3}C_{r-3}-{}_{n-4}C_{r-4})s_{1:x}^2 s_{2:x}$$

$$+4({}_{n-2}C_{r-2}-3\cdot{}_{n-3}C_{r-3}+2\cdot{}_{n-4}C_{r-4})s_{1:x}s_{3:x}$$

$$+3({}_{n-2}C_{r-2}-2\cdot{}_{n-3}C_{r-3}+{}_{n-4}C_{r-4})s_{2:x}^2$$

$$+({}_{n-1}C_{r-1}-7\cdot{}_{n-2}C_{r-2}+12{}_{n-3}C_{r-3}$$

$$-6\cdot{}_{n-4}C_{r-4})s_{4:x}$$

14. Now consider the case where m is any positive integer. Write $x_i^m = y_i$. The operators developed in this chapter will express any power sum with respect to z in terms of the power sums with respect to y, i. e. in terms of $s_{1:y}$, $s_{2:y}$, $s_{3:y}$, ... But obviously $s_{v:y} = s_{mv:x}$ and hence the operators of this chapter will express any symmetric function which is a power sum with respect to z in terms of power sums with respect to x, viz. in terms of $s_{m:x}$, $s_{2m:x}$, $s_{3m:x}$, ... where m is a positive integer. Hence the operators developed in chapters II and III will express any symmetric function of z_i, $i = 1, 2, \ldots, {}_nC_r$, $z_i = \sum^{r:i} x^m$, m a positive integer, in terms of power sums with respect to x. In particular

$$(1)_z = {}_{n-1}C_{r-1} \cdot s_{m \cdot x},$$

$$(2)_z = {}_{n-2}C_{r-2} \cdot s^2_{m:x} + ({}_{n-1}C_{r-1} - {}_{n-2}C_{r-2})\, s_{2m:x}$$

$$(3)_z = {}_{n-3}C_{r-3} \cdot s^3_{m\,x} + 3({}_{n-2}C_{r-2} - {}_{n-3}C_{r-3})\, s_{m:x} s_{2m:x}$$
$$+ ({}_{n-1}C_{r-1} - 3 \cdot {}_{n-2}C_{r-2} + 2 \cdot {}_{n-3}C_{r-3})\, s_{3m:x}$$

15. Consider again the case $m = 1$. $\rho_1 = {}_{n-1}C_{r-1}$, $\rho_2 = {}_{n-2}C_{r-2}$, ... , $\rho_k = {}_{n-k}C_{r-k}$, $k \leqq r$. Then[1]

$$s_{1:z} = \rho_1 s_{1:x},$$

$$s_{2:z} = \rho_2 s^2_{1:x} + (\rho_1 - \rho_2) s_{2:x},$$

$$s_{3:z} = \rho_3 s^3_{1:x} + 3(\rho_2 - \rho_3) s_{1:x} s_{2:x} + (\rho_1 - 3\rho_2 + 2\rho_3) s_{3:x},$$

$$s_{4:z} = \rho_4 s^4_{1:x} + 6(\rho_3 - \rho_4) s^2_{1:x} s_{2:x}$$

[1]Notation suggested in Editorial, *Annals of Mathematical Statistics*, 1 (1930), page 104.

$$+4(\rho_2-3\rho_3+2\rho_4)s_{1:x}s_{3:x}+3(\rho_2-2\rho_3+\rho_4)s_{2:x}^2$$

$$+(\rho_1-7\rho_2+12\rho_3-6\rho_4)s_{4:x},$$

etc.

The question as to whether the coefficients in the above expressions follow any simple law now arises. Instead of $\rho_k = {}_{n-k}C_{r-k}$, $k \leq r$, write $\rho^k = {}_{n-k}C_{r-k}$, $k \leq r$. Let

$$P_1(\rho) = \rho,$$

$$P_2(\rho) = \rho - \rho^2,$$

$$P_3(\rho) = \rho - 3\rho^2 + 2\rho^3,$$

$$P_4(\rho) = \rho - 7\rho^2 + 12\rho^3 - 6\rho^4,$$

etc.

Further, let P_t be the expression obtained from $P_t(\rho)$ by going back to subscripts instead of exponents. Then

$$P_1 = \rho_1,$$

$$P_2 = \rho_1 - \rho_2,$$

$$P_3 = \rho_1 - 3\rho_2 + 2\rho_3,$$

$$P = \rho_1 - 7\rho_2 + 12\rho_3 - 6\rho_4,$$

etc.

The expressions for $s_{1:x}$, $s_{2:x}$, ... may now be written:

$$s_{1:z} = P_1 s_{1:x}\,,$$

$$s_{2:z} = P_1^2 s_{1:x}^2 + P_2 s_{2:x}\,,$$

$$s_{3:z} = P_1^3 s_{1:x}^3 + 3P_1 P_2 s_{1:x} s_{2:x} + P_3 s_{3:x}\,,$$

$$s_{4:z} = P_1^4 s_{1:x}^4 + 6P_1^2 P_2 s_{1:x}^2 s_{2:x} + 4P_1 P_3 s_{1:x} s_{3:x}$$
$$+ 3P_2^2 s_{2:x}^2 + P_4 s_{4:x}\,,$$

etc.

where, of course, $P_r P_s P_t \cdot \cdot \cdot \cdot$ is to be found by multiplying $P_r(\rho)\, P_s(\rho)\, P_t(\rho) \cdot \cdot \cdot \cdot$ and then changing the exponents in the result into subscripts. e. g. To find P_2^2 first find $P_2^2(\rho) = (\rho - \rho^2)^2 = \rho^2 - 2\rho^3 + \rho^4$ and then change the exponents into subscripts, obtaining $P_2^2 = \rho_2 - 2\rho_3 + \rho_4$.

One further step is necessary in order to emphasize the law for the formation of these expressions for $s_{1.z}$, $s_{2.z}$, ...

They may be written in the form

$$s_{1:z} = 1!\left(\frac{P_1 s_{1:x}}{1!}\right),$$

$$s_{2:z} = 2!\left(\frac{P_1^2 s_{1:x}^2}{2!} + \frac{P_2 s_{2:x}}{2!}\right),$$

$$s_{3:z} = 3!\left(\frac{P_1^3 s_{1:x}^3}{3!} + \frac{P_1 P_2 s_{1:x} s_{2:x}}{1!\,2!} + \frac{P_3 s_{3:x}}{3!}\right),$$

$$s_{4:z} = 4!\left(\frac{P_1^4 s_{1:x}^4}{4!} + \frac{P_1^2 P_2 s_{1:x}^2 s_{2:x}}{2!\,1!\,2!}\right.$$

$$\left. + \frac{P_1 P_3 s_{1:x} s_{3:x}}{1!\;3!} + \frac{P_2^2 s_{2:x}^2}{2!(2!)^2} + \frac{P_4 s_{4:x}}{4!}\right)$$

$$s_{t:z} = t!\sum \frac{p_i^I p_j^J p_k^K \cdots s_i^I s_j^J s_k^K \cdots}{(i!)^I (j!)^J (k!)^K \cdots I! J! K! \cdots}$$

After computing by the direct method the first eight moments, under the assumption that $s_{1:x} = M_x = 0$, an article[1] which appeared in the *Annals of Mathematical Statistics* gives the following law for the formation of the functions $P_t(\rho)$ for $t = 1, 2, \cdots, 8$: If $c_{i:t}$ is the coefficient of ρ^i in the expression for $P_t(\rho)$, then

$$c_{i:t} = i c_{i:t-1} - (i-1) c_{i-1:t-1}\,.$$

This is equivalent to saying that

$$P_t(\rho) = \sum_{m=0}^{t-1} \left[(m+1) c_{m+1:t-1} - m c_{m\cdot t-1} \right] \rho^{m+1}.$$

That this law holds for all values of $t = 1, 2, \ldots$ is now easily established For if it be assumed that this law holds for the expression for $(t-1)_z$ in terms of the power sums with respect to x, then it holds also for $(t)_z$ because the operators $D_1, D_2, \cdots, D_t$ and the equivalent operators in terms of $d_1, d_2, \cdots, d_t$ will express $(t)_z$ in terms of the power sums with respect to z and of weight less than t. And the coefficients of the terms in the expression for $(t)_z$ are seen to depend only on the coefficients of these power sums of weight less than t. e. g. Suppose the law holds for $t = 1, 2$. Let

$$(3)_z = Q_1 s_{1:x}^3 + Q_2 s_{1:x} s_{2:x} + Q_3 s_{3:x}\,.$$

[1]Editorial, *Annals of Mathematical Statistics*, 1 (1930), page 107.

Operate on the left with D_1 and on the right with d_1 .

$$3(2)_{x'} = 3Q_1 s_{1\cdot x}^2 + Q_2 s_{2:x} ,$$ hence

$$Q_1 = P_1^3, \quad Q_2 = 3P_1 P_2 .$$

Operate on the left with $6D_3$ and on the right with $(d_1^3 + 6d_1 d_2 + 6d_3)$. Then

$$6P_1 = 6(Q_1 + Q_2 + Q_3),$$ therefore

$$Q_3 = P_1 - Q_1 - Q_2 ,$$

$$= P_1 - P_1^3 - 3P_1 P_2$$

But

$$P_1(\rho) - P_1^3(\rho) - 3P_1(\rho) P_2(\rho) = \rho - \rho^3 - 3\rho(\rho - \rho^2)$$

$$= \rho - 3\rho^2 + 2\rho^3$$

$$= P_3(\rho).$$

Hence

$$Q_3 = P_3$$

16. Consider the functions $P_t(\rho)$, $t = 1, 2, \cdots, 10, \cdots$

$$P_1(\rho) = \rho,$$

$$P_2(\rho) = \rho - \rho^2,$$

$$P_3(\rho) = \rho - 3\rho^2 + 2\rho^3,$$

$$P_4(\rho) = \rho - 7\rho^2 + 12\rho^3 - 6\rho^4,$$

$$P_5(\rho) = \rho - 15\rho^2 + 50\rho^3 - 60\rho^4 + 24\rho^5,$$

$$P_6(\rho) = \rho - 31\rho^2 + 180\rho^3 - 390\rho^4 + 360\rho^5 - 120\rho^6,$$

$$P_7(\rho) = \rho - 63\rho^2 + 602\rho^3 - 2100\rho^4$$

$$+ 336\rho^5 - 2520\rho^6 + 720\rho^7,$$

$$P_8(\rho) = \rho - 127\rho^2 + 1932\rho^3 - 10206\rho^4$$

$$+ 25200\rho^5 - 31920\rho^6 + 20160\rho^7 - 5040\rho^8,$$

$$P_9(\rho) = \rho - 255\rho^2 + 6050\rho^3 - 46620\rho^4 + 166824\rho^5$$

$$- 317520\rho^6 + 332640\rho^7 - 181440\rho^8 + 40320\rho^9,$$

$$P_{10}(\rho) = \rho - 511\rho^2 + 18860\rho^3 - 204630\rho^4$$

$$+ 1020600\rho^5 - 2739240\rho^6 + 3329424\rho^7$$

$$- 3780000\rho^8 + 1814400\rho^9 + 362880\rho^{10}.$$

Those who are familiar with the calculus of finite differences will recognize the coefficients in the above expressions, neglecting their signs, as the numbers appearing in the table of values of

$$\Delta^m 1^n = (\Delta^m x^n)_{x=1}$$

If $u(x)$ and $v(x)$ are functions of x then

$$\Delta^n u(x) . v(x) = v(x) \Delta^n u(x) + {}_nC_1 \cdot \Delta v(x) \Delta^{n-1} u(x+1)$$

$$+ {}_nC_2 \cdot \Delta^2 v(x) \cdot \Delta^{n-2} u(x+2) + \cdot \cdot \cdot$$

Now $x^n = x \cdot x^{n-1}$. Hence, letting $v(x) = x$ and $u(x) = x^{n-1}$, $\Delta^m x^n = \Delta^m x \cdot x^{n-1} = x \Delta^m x^{n-1}$

$+ m \Delta^{m-1}(x+1)^{n-1}$ and all the other terms vanish. Also $(x+1)^{n-1} = E x^{n-1} = (1+\Delta) x^{n-1}$. Therefore

$$\Delta^m x^n = x \Delta^m x^{n-1} + m \Delta^{m-1}(1+\Delta) x^{n-1}$$

$$= x \Delta^m x^{n-1} + m(\Delta^m x^{n-1} + \Delta^{m-1} x^{n-1}).$$

It is now possible to write

$$P_t(\rho) = \sum_{m=0}^{t-1} (-1)^m (\Delta^m 1^{t-1}) \cdot \rho^{m+1}.$$

To show that this law is equivalent to the law given above, viz:

$$P_t(\rho) = \sum_{m=0}^{t-1} \left[(m+1) c_{m+1:\,t-1} - m c_{m:\,t-1} \right] \rho^{m+1}$$

assume they are equivalent for $P_t(\rho)$ and show that they

are then equivalent for $P_{t+1}(\rho)$ That is, assume

$$\sum_{m=0}^{t-1} (-1)^m \left[(m+1)\Delta^m 1^{t-2} + m\Delta^{m-1} 1^{t-2} \right] \rho^{m+1}$$

$$= \sum_{m=0}^{t-1} \left[(m+1)c_{m+1:t-1} - mc_{m:t-1} \right] \rho^{m+1}$$

Then
$(1)^m \Delta^m 1^{t-2} = c_{m+1:t-1}$ and $(-1)^{m-1} \Delta^{m-1} 1^{t-2}$
$c_{m\ t-1}$ For $P_{t+1}(\rho)$ the two laws are equivalent if

$$\sum_{m=0}^{t} (-1)^m \left[(m+1)\Delta^m 1^{t-1} + m\Delta^{m-1} 1^{t-1} \right] \rho^{m+1}$$

$$= \sum_{m=0}^{t} \left[(m+1)c_{m+1:t} - mc_{m:t} \right] \rho^{m+1}$$

But this is true since if $c_{m+1:t-1} = (-1)^m \Delta^m 1^{t-2}$
then

$$\begin{aligned} c_{m+1:t} &= (m+1)c_{m+1:t-1} - mc_{m:t-1} \\ &= (-1)^m \left[(m+1)\Delta^m 1^{t-2} + m\Delta^{m-1} 1^{t-2} \right] \\ &= (-1)^m \Delta^m 1^{t-1}. \end{aligned}$$

Similarly, since $c_{m:t-1} = (-1)^{m-1} \Delta^{m-1} 1^{t-2}$, then

$$c_{m:t} = (-1)^{m-1} \Delta^{m-1} 1^{t-1}.$$

17. Since $\Delta^m 1^n = \Delta^m (1+\Delta) 0^n$, it is possible to write

$$P_t(\rho) = \sum_{m=0}^{t-1} (-1)^m \Delta^m 1^{t-1} \cdot \rho^{m+1}$$
$$= \sum_{m=0}^{t-1} (-1)^m \Delta^m (1+\Delta) x^{t-1}\Big]_{x=0} \rho^{m+1}$$

The latter expression on the right suggests that $P_t(\rho)$ may be expressed as a function of ρ and x with x set equal to zero for each particular value of t. Suppose that $F_t(x,\rho)_{x=0}$ is such a function. Obviously F can be neither a polynomial in x nor a rational function of any kind in x; for setting x equal to zero would show that F would have the same value for all values of t. The nature of the expression suggests that x enters F only as a variable with respect to which differentiation is to be carried out, x then being set equal to zero. There are two main reasons for this assumption. First of all, since x enters the difference expression only as a variable with respect to which differencing is performed, x being set equal to zero after each differencing, the guess is that x enters F only as a variable with respect to which differentiation is to be carried out, x being set equal to zero after each differentiation. Besides this there is the intimate relation between Δ and d/dx. For instance, $1+\Delta = e^{d/dx}$, $d/dx = \log(1+\Delta)$ and hence Δ^n can be replaced by a function of the n'th degree in d/dx and vice versa. Further, since the difference expression contains Δ^t it is reasonable to try to express F as a function involving d^t/dx^t. Now let $F_t(x,\rho)_{x=0} = (d^t/dx^t).\phi(x,\rho)\Big]_{x=0}$. Since t differentiations, none of which are to give results identically zero, are to be carried out then ϕ cannot be a rational function of x. Also functions which involve the possibility of the derivative being infinite are excluded. Hence try a transcendental function of x and ρ. The exponential function will not satisfy the conditions. Try

$\phi(x,\rho) = \log f(x,\rho)$. And again f cannot be a rational function of x. Suppose f is an exponential function of x, say $f(x,\rho) = R(\rho, e^x)$. Then

$$P_t(\rho) = \frac{d^t}{dx^t} \cdot \log R(\rho, e^x)\Big]_{x=0}$$

The simplest case would be $R(\rho, e^x) = \rho e^x$. But this does not satisfy $P_1(\rho) = \rho$. Nor does $R(\rho, e^x) = \rho e^x + \rho$; nor $R(\rho, e^x) = \rho e^x - \rho$. But $R(\rho, e^x) = \rho e^x + 1 - \rho$ does satisfy the conditions since it has been shown[1] that

$$\frac{d^t}{dx^t} \cdot \log(\rho e^x + 1 - \rho)\Big]_{x=0}$$

satisfies the law $c_{i:t} = ic_{i:t-1} - (i-1)c_{i-1:t-1}$, where $c_{i:t}$ is the coefficient of ρ^i in $P_t(\rho)$.

Hence $P_t(\rho)$ can be written in the three equivalent forms for all values of t:

$$P_t(\rho) = \sum_{m=0}^{t-1} (-1)^m (\Delta^m 1^{t-1}) \cdot \rho^{m+1}$$

$$P_t(\rho) = \sum_{m=0}^{t-1} \Big[(m+1)c_{m+1:t-1} - mc_{m:t-1} \Big] \rho^{m+1}.$$

$$P_t(\rho) = \frac{d^t}{dx^t} \log (\rho e^x + 1 - \rho)\Big]_{x=0}$$

[1] Editorial, *Annals of Mathematical Statistics,* 1 (1930), pages 107, 108 Also see remark on "Sampling Polynomials," page 120.

A. L. O'Toole

FUNDAMENTAL FORMULAS FOR THE DOOLITTLE METHOD, USING ZERO-ORDER CORRELATION COEFFICIENTS

By

HAROLD D. GRIFFIN
Dean of Crescent College, Eureka Springs, Arkansas

So far as the writer has been able to determine, fundamental formulas for the Doolittle method as applied to the solution of normal linear equations expressed in correlation coefficients have never before been developed. Because of their peculiar telescoping qualities, the writer has termed them "endothetic formulas." Perhaps the best way to judge the respective merits of three methods of solving simultaneous linear equations to obtain the coefficient of partial regression (the β's)—determinants, Kelley's partial regression method,[1] and Doolittle's direct substitution method[2]—is to compare the formulas by which each might be expressed.

---

[1]Kelley, T. L. Chart to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations. 1st ed. School of Education, Special Monograph No. 1. Palo Alto: Stanford University Publications, 1921.

[2]Wallace, H. A., and Snedecor, G. W. Correlation and Machine Calculation. 1st ed. Official Publ. Vol. 23, No. 35. Ames: Iowa State College of Agriculture, 1925.

THREE-VARIABLE FORMULAS

Determinants

$$\beta_{02.1} = \frac{r_{02} - r_{01} r_{12}}{1 - r_{12}^2} \qquad \beta_{01.2} = \frac{r_{01} - r_{02} r_{12}}{1 - r_{12}^2}$$

Kelley's

$$\beta_{02.1} = \frac{r_{02} - r_{01} r_{12}}{1 - r_{12}^2} \qquad \beta_{01.2} = \frac{r_{01} - r_{02} r_{12}}{1 - r_{12}^2}$$

Doolittle's

$$\beta_{02.1} = \frac{r_{02} - r_{01} r_{12}}{1 - r_{12}^2} \qquad \beta_{01.2} = r_{01} - r_{12} \beta_{02.1}$$

OPERATIONS REQUIRED IN SOLVING A THREE-VARIABLE PROBLEM

	Determinants	Kelley's	Doolittle's
Consulting tables	1	1	1
Adding	0	0	0
Subtracting	2	2	2
Multiplying	2	2	2
Dividing	2	2	1

In a three-variable problem the Doolittle method has but a very slight advantage over the Determinant method and the method used by Kelley in his *Chart*.

FOUR-VARIABLE FORMULAS

Determinants

$$\beta_{03.12} = \frac{r_{03}(1-r_{12}^2) - r_{01}r_{13} - r_{02}r_{23} + r_{12}(r_{01}r_{23} + r_{02}r_{13})}{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}$$

$$\beta_{02.13} = \frac{r_{02}(1-r_{13}^2) - r_{01}r_{12} - r_{03}r_{23} + r_{13}(r_{01}r_{23} + r_{03}r_{12})}{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}$$

$$\beta_{01.23} = \frac{r_{01}(1-r_{23}^2) - r_{02}r_{12} - r_{03}r_{13} + r_{23}(r_{02}r_{13} + r_{03}r_{12})}{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}$$

Kelley's

$$\beta_{03.12} = \frac{\dfrac{r_{03} - r_{01}r_{13}}{1-r_{13}^2} - \dfrac{r_{02} - r_{01}r_{12}}{1-r_{12}^2} \times \dfrac{r_{23} - r_{12}r_{13}}{1-r_{13}^2}}{1 - \dfrac{r_{23} - r_{12}r_{13}}{1-r_{12}^2} \times \dfrac{r_{23} - r_{12}r_{13}}{1-r_{13}^2}}$$

$$\beta_{02.13} = \frac{\dfrac{r_{02} - r_{01}r_{12}}{1-r_{12}^2} - \dfrac{r_{03} - r_{01}r_{13}}{1-r_{13}^2} \times \dfrac{r_{23} - r_{12}r_{13}}{1-r_{12}^2}}{1 - \dfrac{r_{23} - r_{12}r_{13}}{1-r_{13}^2} \times \dfrac{r_{23} - r_{12}r_{13}}{1-r_{12}^2}}$$

$$\beta_{01.23} = \frac{\dfrac{r_{01} - r_{02}r_{12}}{1-r_{12}^2} - \dfrac{r_{03} - r_{02}r_{23}}{1-r_{23}^2} \times \dfrac{r_{13} - r_{12}r_{23}}{1-r_{12}^2}}{1 - \dfrac{r_{13} - r_{12}r_{23}}{1-r_{23}^2} \times \dfrac{r_{13} - r_{12}r_{23}}{1-r_{12}^2}}$$

Doolittle's

$$\beta_{03.12} = \frac{r_{03} - r_{01} r_{13} - \dfrac{r_{02} - r_{01} r_{12}}{1 - r_{12}^2} \times (r_{23} - r_{12} r_{13})}{1 - r_{13}^2 - \dfrac{r_{23} - r_{12} r_{13}}{1 - r_{12}^2} \times (r_{23} - r_{12} r_{13})}$$

$$\beta_{02.13} = \frac{r_{02} - r_{01} r_{12}}{1 - r_{12}^2} - \frac{r_{23} - r_{12} r_{13}}{1 - r_{12}^2} \times \beta_{03.12}$$

$$\beta_{01.23} = r_{01} - r_{12} \beta_{02.13} - r_{13} \beta_{03.12}$$

OPERATIONS REQUIRED IN SOLVING A FOUR-VARIABLE PROBLEM

	Determinants	Kelley's	Doolittle's
Consulting tables	6	3	2
Adding	12	0	1
Subtracting	4	12	4
Multiplying	18	12	8
Dividing	3	11	3

In a four-variable problem the Doolittle method is seen to have a decided advantage over the other two. An examination and comparison of these fundamental formulas for three and four variables would seem to justify the conclusion that an increasing number of variables would but enhance the manifest superiority of the Doolittle method.

Harold D Griffin

ON A PROPERTY OF THE SEMI-INVARIANTS OF THIELE

By

CECIL C. CRAIG
National Research Fellow

Given a general linear form

$$(1) \qquad a_1 x_1 + a_2 x_2 + \cdots\cdots + a_n x_n$$

of a set of statistical variables, $x_1, x_2, \cdots, x_n$,[1] it is well-known that in case the variables, $x_1, x_2, \cdots\cdots, x_n$ are independent, in the sense of the theory of probability, that the r'th semi-invariant of this form is simply

$$(2) \qquad a_1^r \lambda_r^{(1)} + a_2^r \lambda_r^{(2)} + \cdots\cdots + a_n^r \lambda_r^{(n)},^2$$

in which $\lambda_r^{(i)}$ is the r'th semi-invariant of x_i. This is perhaps the most important and useful property of semi-invariants.

Each semi-invariant is defined as a certain isobaric function of the moments of weight equal to the order of the semi-invariant. The question to which this note is devoted is whether among such isobaric functions, the property given above belongs uniquely to the semi-invariant. This problem is equivalent to another which

[1]There is no loss in generality in supposing the origin so chosen for each x_i that the constant in the form is zero.

[2]Thiele, T. N., Theory of Observations (C. & E. Layton, London, 1903) p. 39.

seems more difficult to state verbally. The r'th semi-invariant L_r of the form (1) is itself found in terms of the semi-invariants, $\lambda_{rst\cdots}$, of the n-way probability function $F(x_1, x_2, \cdots, x_n)$ by means of a symbolic multinomial expansion. Now in order that the above property may hold generally it is necessary and sufficient that the cross-semi-invariants of $F(x_1, x_2, \cdots x_n)$ should vanish if $x_1, x_2, \cdots, x_n$ are independent; that is, that each $\lambda_{rst\cdots}$ in which at least two of the quantities $r, s, t, \cdots$ are different from zero, should vanish identically. Now are semi-invariants the only such functions of moments, whose "cross" members behave in this way?

The semi-invariants L_r of the given linear form are defined by

$$e^{L_1 t + \frac{1}{2} L_2 t^2 + \frac{1}{3!} L_3 t^3 + \cdots}$$

$$(3) \qquad = \int_{-\infty, \cdots, -\infty}^{\infty, \cdots, \infty} dF(x_1, x_2, \cdots, x_n) e^{t(\sum_1^n a_i x_i)} \quad {}^{1}$$

which is to be regarded as a formal identity in t. And the semi-invariants of $x_1, x_2, \cdots, x_n$ are given by

$$e^{(\sum_1^n \lambda_i t_i) + \frac{1}{2}(\sum_1^n \lambda_i t_i)^{(2)} + \frac{1}{3!}(\sum_1^n \lambda_i t_i)^{(3)} + \cdots}$$

$$(4) \qquad = \int_{-\infty, \cdots, -\infty}^{\infty, \cdots, \infty} dF(x_1, x_2, \cdots, x_n) e^{(\sum_1^n x_i t_i)}$$

$$= 1 + (\sum_1^n \nu_i t_i) + \frac{1}{2}(\sum_1^n \nu_i t_i)^{(2)} + \frac{1}{3!}(\sum_1^n \nu_i t_i)^{(3)} + \cdots$$

[1] We shall observe the distinction between probability functions and frequency functions suggested by H. Cramér in his important memoir: "On the Composition of Elementary Errors," Skandinavisk Aktuarietidskrift, 1928, p. 13. By a probability function we mean what has been called the cumulative frequency function and thus in the above we are using an n-way Stieltjes integral.

which is also a formal identity in $t_1, t_2, \cdots, t_n$.

The quantities $(\sum_1^n \lambda_i t_i)^{(r)}$ and $(\sum_1^n \nu_i t_i)^{(r)}$ refer to symbolic multinomial expansions, perhaps most easily explained by means of examples. Thus

$$(\sum_1^3 \lambda_i t_i)^{(2)} = \lambda_{200} t_1^2 + \lambda_{020} t_2^2 + \lambda_{002} t_3^2$$
$$+ 2\lambda_{110} t_1 t_2 + 2\lambda_{101} t_1 t_3 + 2\lambda_{011} t_2 t_3 ,$$

and

$$(\sum_1^2 \lambda_i t_i)^{(3)} = \lambda_{30} t_1^3 + \lambda_{03} t_2^3 + 3\lambda_{21} t_1^2 t_2 + 3\lambda_{12} t_1 t_2^2$$

in which $\lambda_{k00\ldots0} = \lambda_k^{(1)}, \lambda_{0k0\ldots0} = \lambda_k^{(2)}, \cdots\cdots,$ in our first used notation, and $\lambda_{110\ldots0}, \lambda_{210\ldots0}$, etc. are cross-semi-invariants of x_1 and x_2.

Then by inspection of (3) and (4) it is evident that

$$(5) \qquad L_k = (\sum_1^n a_i \lambda_i)^{(k)}, \quad k = 1, 2, 3, \cdots\cdots$$

In case the variables $x_1, x_2, \cdots, x_n$ are all independent of each other $F(x_1, x_2, \cdots x_n)$ splits up into the product $F_1(x_1)\ F_2(x_2) \cdots F_n(x_n)$ of the probability functions of the separate variables, L_k becomes equal to the expression (2), and all the cross-semi-invariants in the expansion of the right member of (4) become identically zero. That the vanishing of these cross-semi-invariants is not only a sufficient but is also a necessary condition that L_k assume the value (2) is evident from the absence of any restrictions on $F(x_1, x_2, \cdots, x_n)$ (except that it be an n-way probability function) or on the set $a_1, a_2, \cdots, a_n$.

Now each cross-semi-invariant is expressed as a certain isobaric function of moments, some of them cross-moments. But

in the case of independent variables,

$$\nu_{rst\cdots} = \nu_r \, \nu_s \, \nu_t \cdots .$$

and when this is true, the value of each cross-semi-invariant becomes identically zero. To illustrate this and for use in the demonstration that the semi-invariants are the only such functions, let us write out the fourth order semi-invariants of $F(x_1, x_2, x_3, \cdots, x_n)$ in terms of moments. These are obtained by equating coefficients of like terms in

$$(6) \qquad \begin{aligned} \Big(\sum_1^n \lambda_i t_i\Big)^{(4)} = {} & \Big(\sum_1^n \nu_i t_i\Big)^{(4)} - 4\Big(\sum_1^n \nu_i t_i\Big)^{(3)}\Big(\sum_1^n \nu_i t_i\Big) \\ & - 3\Big[\Big(\sum_1^n \nu_i t_i\Big)^{(2)}\Big]^2 + 12\Big(\sum_1^n \nu_i t_i\Big)^{(2)}\Big(\sum_1^n \nu_i t_i\Big)^2 \\ & - 6\Big(\sum_1^n \nu_i t_i\Big)^4 . \end{aligned}$$ [1]

Leaving off superfluous zeros in the subscripts, this gives for example

$$\lambda_{40} = \nu_{40} - 4\nu_{30}\nu_{10} - 3\nu_{20}^2 + 12\nu_{20}\nu_{10}^2 - 6\nu_{10}^4$$

$$\begin{aligned} \lambda_{22} = {} & \nu_{22} - (2\nu_{21}\nu_{01} + 2\nu_{12}\nu_{10}) - (\nu_{20}\nu_{02} + 2\nu_{11}^2) \\ & + (2\nu_{20}\nu_{01}^2 + 2\nu_{02}\nu_{10}^2 + 8\nu_{11}\nu_{10}\nu_{01}) - 6\nu_{10}^2\nu_{01}^2 . \end{aligned}$$

If in the value of λ_{22} we set $\nu_{22} = \nu_{20}\nu_{02}$, $\nu_{21} = \nu_{20}\nu_{01}$, etc., then $\lambda_{22} \equiv 0$ as it was already known must happen.

For the sake of simplicity let us suppose, at first, that the component variables in (1) are all "equal," that is, that $F(x_1, x_2, \cdots\cdots, x_n) \equiv F(x, x, \cdots\cdots, x)$. In the case of

[1]The general formula giving semi-invariants in terms of moments is to be found in several places. See e. g., C. Jordan, Statistique Mathématique (Gauthier-Villars, Paris, 1927), p. 41. For an elementary derivation and also for an extended example of the use of semi-invariants of a correlation function of several variables see the author's "An Application of Thiele's Semi-invariants to the Sampling Problem," Metron, Vol. VII, No. 4 (1928), pp. 3-74.

independence among $x_1, x_2, \cdots, x_n$ we can write also $F_1(x_1) = F_2(x_2) = \cdots\cdots = F_n(x_n) = F(x)$. An equivalent assumption is that all moments and hence all semi-invariants of the same *type* of $F(x_1, x_2, \cdots, x_n)$ are equal. (Moments of the same type are all those with the same combination of digits in their subscripts.) Then the expressions for all the semi-invariants of the fourth order of $F(x_1, x_2, \cdots, x_n)$ are equivalent to the following:

$$
\begin{aligned}
&\lambda_{40} = \nu_{40} - 4\nu_{30}\nu_{10} - 3\nu_{20}^2 + 12\nu_{20}\nu_{10}^2 - 6\nu_{10}^4 \\
&\lambda_{31} = \nu_{31} - (\nu_{30}\nu_{10} + 3\nu_{21}\nu_{10}) - 3\nu_{20}\nu_{11} + (6\nu_{20}\nu_{10}^2 + 6\nu_{11}\nu_{10}^2) - 6\nu_{10}^4 \\
(7)\ &\lambda_{22} = \nu_{22} - 4\nu_{21}\nu_{10} - (\nu_{20}^2 + 2\nu_{11}^2) + (4\nu_{20}\nu_{10}^2 + 8\nu_{11}\nu_{10}^2) - 6\nu_{10}^4 \\
&\lambda_{211} = \nu_{211} - (2\nu_{21}\nu_{10} + 2\nu_{111}\nu_{10}) - (\nu_{20}\nu_{11} + 2\nu_{11}^2) + (2\nu_{20}\nu_{10}^2 + 10\nu_{11}\nu_{10}^2) - 6\nu_{10}^4 \\
&\lambda_{1111} = \nu_{1111} - 4\nu_{111}\nu_{10} - 3\nu_{11}^2 + 12\nu_{11}\nu_{10}^2 - 6\nu_{10}^4
\end{aligned}
$$

Now, our general isobaric function of the moments of weight four can be written

$$
(8)\quad
\begin{aligned}
f(z_1, z_2, \cdots, z_5) = {} & z_1\Big(\sum_1^n \nu_i t_i\Big)^{(4)} - 4z_2\Big(\sum_1^n \nu_i t_i\Big)^{(3)}\Big(\sum_1^n \nu_i t_i\Big) \\
& - 3z_3\Big[\Big(\sum_1^n \nu_i t_i\Big)^{(2)}\Big]^2 + 12z_4\Big(\sum_1^n \nu_i t_i\Big)^{(2)}\Big(\sum_1^n \nu_i t_i\Big)^2 - 6z_5\Big(\sum_1^n \nu_i t_i\Big)^4
\end{aligned}
$$

And in our special case of equal component variables $x_1, x_2, \cdots, x_n$ our problem is to determine for what sets of values of $z_1, z_2, \cdots, z_5$ the coefficients of $t_1^3 t_2$, $t_1^2 t_2^2$, $t_1^2 t_2 t_3$ and $t_1 t_2 t_3 t_4$ in the right member of (8) vanish identically if $x_1, x_2, \cdots, x_n$ are independent.

By comparison with (7) it is seen that this gives four linear equations with which to determine the five unknowns. But we

can add a fifth equation by stating that the coefficient of t_1^4 is in general a parameter which in the case of independence is a function of $F(x)$ and $z_1, z_2, \cdots, z_5$, which we shall designate by ξ_4. Then we have for the determination of z_1:

$$(9)\; z_1 = \frac{\begin{vmatrix} \xi_4 & -4\nu_3\nu_1 & -3\nu_2^2 & 12\nu_2\nu_1^2 & -6\nu_1^4 \\ 0 & -(\nu_3\nu_1+3\nu_2\nu_1^2) & -3\nu_2\nu_1^2 & 6\nu_2\nu_1^2+6\nu_1^4 & -6\nu_1^4 \\ 0 & -4\nu_2\nu_1^2 & -(\nu_2^2+2\nu_1^4) & 4\nu_2\nu_1^2+8\nu_1^4 & -6\nu_1^4 \\ 0 & -(2\nu_2\nu_1^2+2\nu_1^4) & -(\nu_2\nu_1^2+2\nu_1^4) & 2\nu_2\nu_1^2+10\nu_1^4 & -6\nu_1^4 \\ 0 & -4\nu_1^4 & -3\nu_1^4 & 12\nu_1^4 & -6\nu_1^4 \end{vmatrix}}{\begin{vmatrix} \nu_4 & -\nu_3\nu_1 & -3\nu_2^2 & 12\nu_2\nu_1^2 & -6\nu_1^4 \\ \nu_3\nu_1 & -(\nu_3\nu_1+3\nu_2\nu_1^2) & -3\nu_2\nu_1^2 & 6\nu_2\nu_1^2+6\nu_1^4 & -6\nu_1^4 \\ \nu_2^2 & -4\nu_2\nu_1^2 & -(\nu_2^2+2\nu_1^4) & 4\nu_2\nu_1^2+8\nu_1^4 & -6\nu_1^4 \\ \nu_2\nu_1^2 & -(2\nu_2\nu_1^2+2\nu_1^4) & -(\nu_2\nu_1^2+2\nu_1^4) & 2\nu_2\nu_1^2+10\nu_1^4 & -6\nu_1^4 \\ \nu_1^4 & -4\nu_1^4 & -3\nu_1^4 & 12\nu_1^4 & -6\nu_1^4 \end{vmatrix}}$$

By adding each of the four other columns to the first column in the denominator, we have at once in view of (7),

$$z_1 = \xi_4 / \lambda_4$$

unless the identical first minor of numerator and denominator vanishes. But this can happen only if there is linear dependence between the corresponding elements in the four rows of this minor which in turn can happen only if there is a linear relation between the quantities $\nu_3\nu_1$, ν_2^2, $\nu_2\nu_1^2$, and ν_1^4. (Such a linear dependence would exist if the second or third semi-invariant of $F(x)$ is zero.)

Moreover, it is readily seen that we get $z_1 = z_2 = \cdots = z_5 = \dfrac{\xi_4}{\lambda_4}$ (Of course we suppose $\lambda_4 \neq 0$ and moreover $\xi_4 = 0$ could hold only for some $F(x)$'s)

If we no longer suppose the components $x_1, x_2, \cdots, x_n$ "equal" in the sense defined above, the quantities in (7) may be replaced by summations of all terms of the same type or summations of all products of terms which are coefficients of similar

terms in t_i 's. Thus in place of $\lambda_{40}, \nu_{40}, \nu_{30}, \nu_{10}$ in the first equation, and λ_{31} and $\nu_{30}\, \nu_{01}$ in the second we now write,

$$\Sigma\, \lambda_{40} = \lambda_{40} + \lambda_{04} + \lambda_{004} + \cdots$$

$$\Sigma\, \nu_{40} = \nu_{40} + \nu_{04} + \nu_{004} + \cdots$$

$$\Sigma\, \nu_{30}\nu_{10} = \nu_{30}\nu_{10} + \nu_{03}\nu_{01} + \nu_{003}\nu_{001} + \cdots$$

$$\Sigma\, \lambda_{31} = \lambda_{31} + \lambda_{13} + \lambda_{031} + \lambda_{013} + \cdots$$

$$\Sigma\, \nu_{30}\nu_{01} = \nu_{30}\nu_{01} + \nu_{03}\nu_{10} + \nu_{030}\nu_{001} + \nu_{003}\nu_{010} + \cdots,$$

respectively. But otherwise our argument will be the same and lead to the same conclusion.

It is obvious that the argument for weight four is perfectly general and thus that the same kind of conclusions hold for any weight. We conclude that the semi-invariants are the only isobaric functions of the moments of a set of n variables which have the properties described in the first two paragraphs independent of the probability or frequency functions of those variables.

But if when the variables are independent the probability function of each one is such that there is an isobaric relation among the moments of order lower than k, the same for each variable, then there are other isobaric functions of order k and higher which enjoy the property of semi-invariants in question. And it will be shown that the only isobaric relations among the moments of order $< k$, mentioned above, which lead to the new isobaric functions of this type of order $\geq k$, are obtained by setting semi-invariants of order $< k$, equal to zero.

Let us return to the case in which the weight is four. Then if $\lambda_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^{\,3} = 0$, the minor D_{11} of our denominator D vanishes, and so, of course, does the corresponding minor in the numerator. Then as a matter of fact there is a double infinity of the sought isobaric functions of weight four.

Some of them are given by the following sets of values of the z's.

z_1	z_2	z_3	z_4	z_5
5	2	5	2	1
6	3	6	3	2
9	3	9	3	1

as may be verified by actual computation.

Now we also have[1]

$$\lambda_4 = \nu_4 - \lambda_1 \nu_3 - 3\lambda_2 \nu_2 - 3\lambda_3 \nu_1$$

from which we can write in place of (8)

$$(10)\qquad \begin{aligned} f(y_1,\cdots,y_4) = {} & y_1\Big(\sum_1^n \nu_i t_i\Big)^{(4)} - y_2\Big(\sum_1^n \lambda_i t_i\Big)\Big(\sum_1^n \nu_i t_i\Big)^{(3)} \\ & -3y_3\Big(\sum_1^n \lambda_i t_i\Big)^{(2)}\Big(\sum_1^n \nu_i t_i\Big)^{(2)} - 3y_4\Big(\sum_1^n \lambda_i t_i\Big)^{(3)}\Big(\sum_1^n \nu_i t_i\Big) \end{aligned}$$

in which we can seek to find sets of values of $y_1, \cdots\cdots, y_4$ so that the coefficients of $t_1^3 t_2$, $t_1^2 t_2^2$, $t_2^2 t_3 t_4$ and $t_1 t_2 t_3 t_4$ will vanish when the x's are independent. This will give us four homogeneous linear equations in which the determinant of the coefficients vanishes identically since $y_1 = y_2 = y_3 = y_4 = 1$ is a solution. Addition of the second, third and fourth columns to the first gives a new first column of zeros. But if, say, $\lambda_3 = 0$, in addition to λ_{21}, and λ_{111} which already vanish if the x's are independent, then the elements of the fourth column are all zeros also, and our determinant is of rank not greater than two. But since the solution of the set of equations arising from (10) is equivalent to that arising from (8), the minor D_{11}, of D in (9)

[1]Thiele, T. N., loc. cit., p. 25.

must vanish in case $\lambda_3 = 0$.

But since $z_1 = z_2 = \cdots = z_5 = 1$ is a solution of the equations (8), it is easy to see that if in D_{II}, the sum of the last three columns be added to the first column, the resulting first column will be identical, though opposite in sign with the last four elements of the first coulmn of D. Let us indicate the new D_{II} by D_{II}'.

Now there is a linear dependence between the elements of the rows of D_{II}. In fact the elements of the first row minus three times the corresponding elements of the third plus twice the corresponding elements of the fourth ($\lambda_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3$) must give zero for each element. For suppose there exists another such linear relationship between rows. This linear relationship must hold between the corresponding elements of the first column of D_{II}', and we have a new isobaric relation between the moments of x. But a probability function $F(x)$can always be found in which

$$\nu_3\nu_1 - 3\nu_2\nu_1^2 + 2\nu_1^4 = \lambda_3\nu_1 = 0 \tag{11}$$

holds and the other relation does not. But for the $F(x)$'s in which (11) holds D_{II}' must vanish, and thus the relation between columns must be that given by (11).

Thus D_{II} contains as factors λ_3, λ_2 and λ_1. That it contains no others can easily be verified directly.

The cases of weights two, three, and four are easily handled directly throughout. If the weight is now k greater than four, our argument readily generalizes. The equations now arising from the relation corresponding to (10) are now greater in number than the unknowns $y_1, y_2, \cdots, y_k$, but it is obvious that the matrix of the coefficients is of rank not greater than $k-2$. And it follows just as before that $\lambda_{k-1}, \lambda_{k-2}, \cdots, \lambda_1$, are all factors of the new D_{II}.

The argument above which shows for the weight four, that

λ_3 is a factor of D_{11} does not show that there cannot be other linear relations between the elements of the first column which are also factors of D_{11} It only shows that if there is such a factor, the corresponding linear dependence holds for certain rows of D_{11}

Let us consider the case of weight five The elements of the first column of D are now $\nu_5, \nu_4\nu_1, \nu_3\nu_2, \nu_3\nu_1^2, \nu_2^2\nu_1, \nu_2\nu_1^3$ and ν_1^5 and the elements of the first column of D'_{11} are the last six of these with opposite sign, and they thus correspond to the partitions of 5. We know that one of the two sets of three rows of D_{11}, the second, fourth, and fifth or the third, fifth, and sixth, are connected by the linear relation corresponding to $\nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 = \lambda_3 = 0$ so that λ_3 is at least once a factor of D_{11}. If we suppose that the first set of three rows are so related, does it follow that this same relation holds for the second set? Now it is easy to see that if in the second row ν_1^2 be everywhere substituted for ν_2 the resulting row will be identical with the third and that the same is true of the fourth and fifth rows and of the fifth and sixth. Then if a certain linear relation holds for the first set of three rows, by the substitution of ν_1^2 for ν_2 everywhere in it, it follows that the same relation holds for the second set of three rows also. Thus λ_3 is twice a factor in D_{11} for weight five. We note also that the partitions of 3 (counting 3 as a partition of 3) are twice found with common factors among the partitions of 5, that is, 32, 221, 2111; and 311, 2111, 11111.

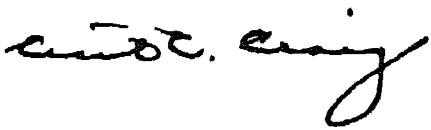
The argument is readily generalized[1] and in case of D_{11} of weight k, each semi-invariant of weight $r < k$ is a factor of D_{11}

[1] The general argument is based on the principle that the second row of D is obtained from the process which gives the first by replacing one factor t_1 by t_2, the third from the first by replacing t_1^2 by t_2^2, the fourth from the first by replacing t_1^2 by $t_2 t_3$, and so on (see (6) and (7)). Thus in the case of weight six, to compare the three rows beginning with $\nu_3^2, \nu_3\nu_2\nu_1, \nu_3\nu_1^3$ with the three beginning with $\nu_3\nu_1^3, \nu_2\nu_1^4, \nu_1^6$, we replace the ν_3 in the first set which arises as a coefficient of t_1^3 by ν_1^3 and the two sets of rows become identical

as often as the partitions of r are found with common factors among the partitions of k. (We count r as a partition of r.) Thus for weight four, $D_{11} = \lambda_3 \lambda_2^3 \lambda_1^7$ which gives D_{11} the correct weight sixteen. In case of weight five, $D_{11} = \lambda_4 \lambda_3^2 \lambda_2^4 \lambda_1^{12}$ which again gives D_{11} the correct weight thirty. And it is easy to show by induction that in case of weight k this method gives D_{11} its proper weight. Among the partitions of k are found all the partitions of $k-1$ with a part 1 added to each. Thus each of these adds k to the total weight. For the partition $k-2, 2$, it is seen that the remaining partitions of $k-2$ with the common additional part 2 will be found among the remaining partitions of k and that the remaining partitions of 2 with the common additional part $k-2$ will also be found. Thus this partition contributes the weight k to the total. And sim arly it can be seen that every partition of k contributes k to the total weight of D_{11}, which was to be proved.

Finally, then, we have the additional result that the necessary and sufficient condition that more than one isobaric function of weight k of the moments of the probability variables $x_1, x_2, \cdots, x_n$ exists which has the semi-invariant properties in question, is that the probability functions of $x_1, x_2, \cdots, x_n$ in case of independence are such that for some $r < k$, λ_r vanishes for each of them.

Stanford University,



THE THEORY OF OBSERVATIONS

By

T. N. THIELE

EDITOR'S NOTE

Thiele's "Theory of Observations" constitutes a classic contribution to both mathematical statistical theory and the theory of least squares. Unfortunately, his researches, and in particular his semi-invariant or "half-invariant" theory, have not received the recognition in this country that they deserve. Since, according to importers of books, the "Theory of Observations" is now out of print and copies are rare, the editor has deemed it advisable as a matter of policy to make this work in this way available to the readers of the *Annals*.

This reprint should also be construed as an acknowledgment of our indebtedness to Mr. Arne Fisher for his unswerving endeavors to bring before American statisticians the important contributions of Danish and Scandinavian writers.

CONTENTS.

VI. Laws of Errors expressed by Symmetrical Functions.

VII. Relations between Functional Laws of Errors and Half Invariants.

VIII. Laws of Errors of Functions of Observations.

IX. Free Functions.

X. Adjustment.

XI. Adjustment by Correlates.

XII. Adjustment by Elements.

XIII. Special Auxiliary Methods.

XIV. The Theory of Probability.

XV. The Formal Theory of Probability.

XVI. The Determination of Probabilities a Priori and a Posteriori.

XVII. Mathematical Expectation and its Mean Error.

I. THE LAW OF CAUSALITY.

§ 1 We start with the assumption that *everything that exists, and everything that happens, exists or happens as a necessary consequence of a previous state of things* If a state of things is repeated in every detail, it must lead to exactly the same consequences. Any difference between the results of causes that are in part the same, must be explainable by some difference in the other part of the causes.

This assumption, which may be called the law of causality, cannot be proved, but must be believed; in the same way as we believe the fundamental assumptions of religion, with which it is closely and intimately connected. The law of causality forces itself upon our belief It may be denied in theory, but not in practice. Any person who denies it, will, if he is watchful enough, catch himself constantly asking himself, if no one else, why *this* has happened, and not *that* But in that very question he bears witness to the law of causality. If we are consistently to deny the law of causality, we must repudiate all observation, and particularly all prediction based on past experience, as useless and misleading

If we could imagine for an instant that the same complete combination of causes could have a definite number of different consequences, however small that number might be, and that among these the occurrence of the actual consequence was, in the old sense of the word, accidental, no observation would ever be of any particular value. Scientific observations cannot be reconciled with polytheism. So long as the idea prevailed that the result of a journey depended on whether the power of Njord or that of Skade was the stronger, or that victory or defeat in battle depended on whether Jove had, or had not, listened to Juno's complaints, so long were even scientists obliged to consider it below their dignity to consult observations

But if the law of causality is acknowledged to be an assumption which always holds good, then every observation gives us a revelation which, when correctly appraised and compared with others, teaches us the laws by which God rules the world.

We can judge of the far-reaching consequences it would have, if there were conditions in which the law of causality was not valid at all, by considering the cases in which the effects of the law are more or less veiled.

In inanimate nature the relation of cause and effect is so clear that the effects are determined by observable causes belonging to the condition immediately preceding, so that the problem, within this domain, may be solved by a tabular arrangement of the several observed results according to the causing circumstances, and the transformation of the tables into laws by means of interpolation. When, however, living beings are the object of our observations, the case immediately becomes more complicated.

It is the prerogative of living beings to hide and covertly to transmit the influences received, and we must therefore within this domain look for the influencing causes throughout the whole of the past history. A difference in the construction of a single cell may be the only indication present at the moment of the observation that the cell is a transmitter of the still operative cause, which may date from thousands of years back. In consequence of this the naturalist, the physiologist, the physician, can only quite exceptionally attain the same simple, definite, and complete accordance between the observed causes and their effects, as can be attained by the physicist and the astronomer within their domains.

Within the living world, communities, particularly human ones, form a domain where the conditions of the observations are even more complex and difficult. Living beings hide, but the community deceives. For though it is not in the power of the community either to change one tittle of any really divine law, or to break the bond between cause and effect, yet every community lays down its own laws also. Every community tries to give its law fixity, and to make it operate as a cause; for instance, by passing it off as divine or by threats of punishment, but nevertheless the laws of the community are constantly broken and changed.

Statistical Science which, in the case of communities, represents observations, has therefore a very difficult task; although the observations are so numerous, we are able from them alone to answer only a very few questions in cases where the intellectual weapons of historical and speculative criticism cannot assist in the work, by independently bringing to light the truths which the communities want to conceal, and on the other hand by removing the wrong opinions which these believe in and propagate.

§ 2. An isolated sensation teaches us nothing, for it does not amount to an observation. Observation is a putting together of several results of sensation which are or are supposed to be connected with each other according to the law of causality, so that some represent causes and others their effects.

By virtue of the law of causality we must believe that, in all observations, we get essentially correct and true revelations; the difficulty is, to ask searchingly enough and to understand the answer correctly. In order that an observation may be free from every other assumption or hypothesis than the law of causality, it must include a perfect

description of all the circumstances in the world, at least at the instant preceding that at which the phenomenon is observed But it is clear that this far surpasses what can be done, even in the most important cases. Real observations have a much simpler form. By giving a short statement of the time and place of observation, we refer to what is known of the state of things at the instant; and, of the infinite multiplicity of circumstances connected with the observation we, generally, not only disregard everything which may be supposed to have little or no influence, but we pay attention only to a small selection of circumstances, which we call *essential*, because we expect, in virtue of a special hypothesis concerning the relation of cause and effect, that the observed phenomenon will be effect of these circumstances only

Nay, we are often compelled to disregard certain circumstances as *unessential*, though there is no doubt as to their influencing the phenomenon, and we do this either because we cannot get a sufficient amount of trustworthy information regarding them, or because it would be impracticable to trace out their connection with the effect For instance in statistical observations on mortality, where the age at the time of death can be regarded as the observed phenomenon, we generally mention the sex as an essential circumstance, and often give a general statement as to residence in town or country, or as to occupation. But there are other things as to which we do not get sufficient information: whether the dead person has lived in straitened or in comfortable circumstances, whether he has been more or less exposed to infectious disease, etc.; and we must put up with this, even if it is certain that one or other of these things was the principal cause of death. And analogous cases are frequently met with both in scientific observations and in everyday occurrences

In order to obtain a perfect observation it is necessary, moreover, that our sensations should give us accurate information regarding both the phenomenon and the attendant circumstances, but all our senses may be said to give us merely approximate descriptions of any phenomenon rather than to measure it accurately Even the finest of our senses recognizes no difference which falls short of a certain finite magnitude This lack of accuracy is, moreover, often greatly increased by the use of arbitrary round numbers for the sake of convenience The man who has to measure a race-course, may take into account the odd metres, but certainly not the millimetres, not to mention the microns.

§ 3. *Owing to all this, every actual observation is affected with errors.* Even our best observations are based upon hypothesis, and often even on an hypothesis that is certainly wrong, namely, that only the circumstances which are regarded as essential, influence the phenomenon; and a regard for practicability, expense, and convenience makes us give approximate estimates instead of the sharpest possible determinations.

Now and then the observations are affected also by *gross errors* which, although

not introduced into them on purpose, are yet caused by such carelessness or neglect that they could have been, and ought to have been, avoided. In contradistinction to these we often call the more or less unavoidable errors *accidental*. For accident (or chance) is not, what the word originally meant, and what still often lingers in our ordinary acceptation of it, a capricious power which suffers events to happen without any cause, but only a name for the unknown element, involved in some relation of cause and effect, which prevents us from fully comprehending the connection between them. When we say that it is accidental, whether a die turns up "six" or "three", we only mean that the circumstances connected with the throwing, the fall, and the rolling of the die are so manifold that no man, not even the cleverest juggler and arithmetician united in the same person, can succeed in controlling or calculating them.

In many observations we reject as unessential many circumstances about which we really know more or less. We may be justified in this; but if such a circumstance is of sufficient importance as a cause, and we arrange the observations with special regard to it, we may sometimes observe that the errors of the observations show a regularity which is not found in "accidental" errors. The same may be the case if, in computations dealing with the results of observations, we make a wrong supposition as to the operation of some circumstance. Such errors are generally called *systematic*.

§ 4. It will be found that every applied science, which is well developed, may be divided into two parts, a theoretical (speculative or mathematical) part and an empirical (observational) one. Both are absolutely necessary, and the growth of a science depends very much on their influencing one another and advancing simultaneously. No lasting divergence or subordination of one to the other can be allowed.

The theoretical part of the science deals with what we suppose to be accurate determinations, and the object of its reasonings is the development of the form, connection, and consequences of the hypotheses. But it must change its hypotheses as soon as it is clear that they are at variance with experience and observation.

The empirical side of the science procures and arranges the observations, compares them with the theoretical propositions, and is entitled by means of them to reject, if necessary, the hypotheses of the theory By induction it can deduce laws from the observations. But it must not forget — though it may have a natural inclination to do so — that, as shown above, it is itself founded on hypotheses. The very form of the observation, and especially the selection of the circumstances which are to be considered as essential and taken into account in making the several observations, must not be determined by rule of thumb, or arbitrarily, but must always be guided by theory.

Subject to this it must as a rule be considered best, that the two sides of the science should work somewhat independently of one another, each in its own particular

way. In what follows the empirical side will be treated exclusively, and it will be treated on a general plan, investigating not the particular way in which statistical, chemical, physical, and astronomical observations are made, but the common rules according to which they are all submitted to computation

II. LAWS OF ERRORS.

§ 5. Every observation is supposed to contain information, partly as to the phenomenon in which we are particularly interested, partly as to all the circumstances, connected with it, which are regarded as essential. In comparing several observations, it makes a very great difference, whether such essential circumstances have remained unchanged, or whether one or several of them have changed between one observation and another. The treatment of the former case, that of *repetitions*, is far simpler than that of the latter, and is therefore more particularly the subject of our investigations; nevertheless, we must try to master also the more difficult general case in its simplest forms, which force themselves upon us in most of the empirical sciences.

By *repetitions* then we understand those observations, in which all the essential circumstances remain unchanged, in which therefore the results or phenomena should agree, if all the operative causes had been included among our essential circumstances. Furthermore, we can without hesitation treat as repetitions those observations, in which we assume that no essential circumstance has changed, but do not know for certain that there has been no such change. Strictly speaking, this would furnish an example of observations with systematic errors; but provided there has been no change in the care with which the essential circumstances have been determined or checked, it is permissible to employ the simpler treatment applicable to the case of repetitions This would not however be permissible, if, for instance, the observer during the repetitions has perceived any uncertainty in the records of a circumstance, and therefore paid greater attention to the following repetitions.

§ 6 The special features of the observations, and in particular their degree of accuracy, depend on causes which have been left out as unessential circumstances, or on some overlooked uncertainty in the statement of the essential circumstances. Consequently no speculation can indicate to us the accuracy and particularities of observations. These must be estimated by comparison of the observations with each other, but only in the case of repetitions can this estimate be undertaken directly and without some preliminary work. The phrase *law of errors* is used as a general name for any mathematical expression representing the distribution of the varying results of repetitions.

Laws of actual errors are such as correspond to repetitions actually carried out. But observations yet unmade may also be erroneous, and where we have to speak hypothetically about observations, or have to do with the prediction of results of future repetitions, we are generally obliged to employ the idea of "laws of errors". In order to prevent any misunderstanding we then call this idea "*laws of presumptive errors*". The two kinds of laws of errors cannot generally be quite the same thing. Every variation in the number of repetitions must entail some variations in the corresponding law of errors; and if we compare two laws of actual errors obtained from repetitions of the same kind in equal number, we almost always observe great differences in every detail. In passing from actual repetitions to future repetitions, such differences at least are to be expected. Moreover, whilst any collection of observations, which can at all be regarded as repetitions, will on examination give us its law of actual errors, it is not every series of repetitions that can be used for predictions as to future observations. If, for instance, in repeated measurements of an angle, the results of our first measurements all fell within the first quadrant, while the following repetitions still more frequently, and at last exclusively, fell within the second quadrant, and even commenced to pass into the third, it would evidently be wrong to predict that the future repetitions would repeat the law of actual errors for the totality of these observations. In similar cases the observations must be rejected as bad or misconceived, and no law of presumptive errors can be directly based upon them.

§ 7. Suppose, however, that on comparing repetitions of some observation we have several times determined the law of actual errors in precisely the same way, employing at first small numbers of repetitions, then larger and still larger numbers for each law. If then, on comparing these laws of actual errors with one another, we remark that they become more alike in proportion as the numbers of repetitions grow greater, and that the agreements extend successively to all those details of the law which are not by necessity bound to vary with the number of repetitions, then we cannot have any hesitation in using the law of actual errors, deduced from the largest possible number of repetitions, for predictions concerning future observations, made under essentially the same circumstances.

This, however, is wholly legitimate only, when it is to be expected that, *if we could obtain repetitions in indefinitely increasing numbers, the law of errors would then approach a single definite form, namely the law of presumptive errors* itself, and would not oscillate between several forms, or become altogether or partly indeterminate. (Note the analogy with the difference between converging and oscillating infinite series). We must therefore distinguish between good and bad observations, and only the good ones, that is those which satisfy the above mentioned condition, *the law of large numbers*, yield laws of presumptive errors and afford a basis for prediction.

As we cannot repeat a thing indefinitely often, we can never be quite certain that

a given method of observation may be called good. Nevertheless, we shall always rely on laws of actual errors, deduced from very large numbers of concordant repetitions, as sufficiently accurate approximations to the law of presumptive errors.

And, moreover, the purely hypothetical assumption of the existence of a law of presumptive errors may yield some special criteria for the right behaviour of the laws of actual errors, corresponding to the increasing number of the repetitions, and establish the conditions necessary to justify their use for purposes of prediction.

We must here notice that, when a series of repetitions by such a test proves bad and inapplicable, we shall nevertheless often be able, sometimes by a theoretical criticism of the method, and sometimes by watching the peculiarities in the irregularities of the laws of errors, to find out the reason why the given method of observation is not as good as others, and to change it so that the checks will at least show that it has been improved. In the case mentioned in the preceding paragraph, for instance, the remedy is obvious. The time of observation is there to be reckoned among the essential circumstances

And if we do not attain our object, but should fail in many attempts at throwing light upon some phenomenon by means of good observations, it may be said even at this stage, before we have been made acquainted with the various means that may be employed, and the various forms taken by the laws of errors, that absolute abandonment of the law of large numbers, as quite inapplicable to any given refractory phenomenon, will generally be out of the question. After repeated failures we may for a time give up the whole matter in despair; but even the most thorough sceptic may catch himself speculating on what may be the cause of his failure, and, in doing so, he must acknowledge that the error is never to be looked for in the objective nature of the conditions, but in an insufficient development of the methods employed. From this point of view then the law of large numbers has the character of a belief. There is in all external conditions such a harmony with human thought that we, sooner or later, by the use of due sagacity, particularly with regard to the essential subordinate circumstances of the case, will be able to give the observations such a form that the laws of actual errors, with respect to repetitions in increasing numbers, will show an approach towards a definite form, which may be considered valid as the law of presumptive errors and used for predictions.

§ 8. Four different means of representing the law of errors must be described, and their respective merits considered, namely:

Tabular arrangements,
Curves of Errors,
Functional Laws of Errors,
Symmetric Functions of the Repetitions.

In comparing these means of representing the laws of errors, we must take into

consideration which of them is the easiest to employ, and neither this nor the description of the forms of the laws of errors demands any higher qualification than an elementary knowledge of mathematics. But we must take into account also, how far the different forms are calculated to emphasise the important features of the laws of errors, i. e. those which may be transferred from the laws of actual errors to the laws of presumptive errors. On this single point, certainly, a more thorough knowledge of mathematics would be desirable than that which may be expected from the majority of those students who are obliged to occupy themselves with observations. As the definition of the law of presumptive errors presupposes the determination of limiting values to infinitely numerous approximations, some propositions from the differential calculus would, strictly speaking, be necessary.

III. TABULAR ARRANGEMENTS.

§ 9. In stating the results of all the several repetitions we give the law of errors in its simplest form. Identical results will of course be noted by stating the number of the observations which give them.

The table of errors, when arranged, will state all the various results and the frequency of each of them.

The table of errors is certainly improved, when we include in it the *relative frequencies* of the several results, that is, the ratio which each absolute frequency bears to the total number of repetitions. It must be the *relative* frequencies which, according to the law of large numbers, are, as the number of observations is increased, to approach the constant values of the law of presumptive errors. Long usage gives us a special word to denote this transition in our ideas: *probability* is the relative frequency in a law of presumptive errors, the proportion of the number of coincident results to the total number, on the supposition of infinitely numerous repetitions. There can be no objection to considering the *relative frequency* of the law of actual errors as an approximation to the corresponding *probability* of the law of presumptive errors, and the doubt whether the relative frequency itself is the best approximation that can be got from the results of the given repetitions, is rather of theoretical than practical interest. Compare § 73.

It makes some difference in several other respects — as well as in the one just mentioned — if the phenomenon is such that the results of the repetitions show qualitative differences or only differences of magnitude.

§ 10. In the former case, in which no transition occurs, but where there are such abrupt differences that none of the results are more closely connected with one another than with the rest, the tabular form will be the only possible one, in which the law of errors can

be given This case frequently occurs in statistics and in games of chance, and for this reason the theory of probabilities, which is the form of the theory of observations in which these cases are particularly taken into consideration, demands special attention. All previous authors have begun with it, and made it the basis of the other parts of the science of observation I am of opinion, however, that it is both safer and easier to keep it to the last.

§ 11. If, however, there is such a difference between the results of repetitions, that there is either a continuous transition between them, or that some results are nearer each other than all the rest, there will be ample opportunity to apply mathematical methods; and when the tabular form is retained, we must take care to bring together the results that are near one another. A table of the results of firing at a target may for instance have the following form:

	1 foot to the left	Central	1 foot to the right	Total
1 foot too high ...	3	17	6	26
Central .	13	109	19	141
1 foot too low	4	8	1	13
Total .	20	134	26	180

If here the heading "1 foot to the left" means that the shot has swerved to the left between half a foot and one foot and a half, this will remind us that we cannot give the exact measures in such tables, but are obliged to give them in round numbers The number of results then will not correspond to such as were exactly the same, but disregarding small differences, we gather into each column those that approach nearest to one another, and which all fall within arbitrarily chosen limits.

In the simple case, where the result of the observation can be expressed by a single real number, the arranged table not only takes the extremely simple form of a table of functions with a single argument, but, as we shall see in the following chapters, leads us to the representation of the law of errors by means of curves of errors and functional laws of errors.

It is an obvious course to fix the attention on the two extreme results in the table, and not seldom these alone are given, instead of a law of error, as a sort of index of the exactness of the whole series of repetitions, and as the higher and lower limits of the observed phenomenon. This index of exactness, however, must be rejected as itself too inexact for the purpose, for the oftener the observations are repeated, the farther we must expect the extremes to move from one another; and thus the most valuable series of observations will appear to possess the greatest range of discrepancy

On the other hand, if, in a table arranged according to the magnitude of the values, we select a single middle value, preceded and followed by nearly equal numbers of values, we shall get a quantity which is very well fitted to represent the whole series of repetitions.

If, while we are thus counting the results arranged according to their magnitude, we also take note of these two values with which we respectively (*a*) leave the first sixth part of the total number, and (*b*) enter upon the last sixth part (more exactly we ought to say 16 per ct.), we may consider these two as indicating the limits between great and small deviations. If we state these two values along with the middle one above referred to, we give a serviceable expression for the law of errors, in a way which is very convenient, and although rough, is not to be despised. Why we ought to select just the middle value and the two sixth-part values for this purpose, will appear from the following chapters.

IV. CURVES OF ERRORS.

§ 12. Curves of actual errors of repeated observations, each of which we must be able to express by one real number, are generally constructed as follows. On a straight line as the axis of abscissae, we mark off points corresponding to the observed numerical quantities, and at each of these points we draw an ordinate, proportional to the number of the repetitions which gave the result indicated by the abscissa. We then with a free hand draw the curve of errors through the ends of the ordinates, making it as smooth and regular as possible. For quantities and their corresponding abscissae which, from the nature of the case, *might* have appeared, but do not really appear, among the repetitions, the ordinate will be = 0, or the point of the curve falls on the axis of abscissae. Where this case occurs very frequently, the form of the curves of errors becomes very tortuous, almost discontinuous. If the observation is essentially bound to discontinuous numbers, for instance to integers, this cannot be helped.

§ 13. If the observation is either of necessity or arbitrarily, in spite of some inevitable loss of accuracy, made in round numbers, so that it gives a lower and a higher limit for each observation, a somewhat different construction of the curve of errors ought to be applied, viz. such a one, that the area included between the curve of error, the axis of abscissae, and the ordinates of the limits, is proportional to the frequency of repetitions within these limits. But in this way the curve of errors may depend very much on the degree of accuracy involved in the use of round numbers. This construction of areas can be made by laying down rectangles between the bounding ordinates, or still better, trapezoids with their free sides approximately parallel to the tangents of the curve. If the

limiting round numbers are equidistant, the mean heights of the trapezoids or rectangles are directly proportional to the frequencies of repetition. In this case a preliminary construction of curve-points can be made as in § 12, and may often be used as sufficient.

It is a very common custom, but one not to be recommended, to draw a broken line between the observed points instead of a curve.

§ 14. There can be no doubt that the curve of errors, as a form for the law of errors, has the advantage of perspicuity, and were not the said uncertainty in so many cases a critical drawback, this would perhaps be sufficient. Moreover, it is in practice quite possible, and not very difficult, to pass from the curve of actual errors to one which may hold good for presumptive errors; though, certainly, this transition cannot be founded upon any positive theory, but depends on skill, which may be acquired by working at good examples, but must be practised judiciously.

According to the law of large numbers we must expect that, when we draw curves of actual errors according to relative frequency, for a numerous series of repetitions, first based upon small numbers, afterwards redrawn every time as we get more and more repetitions, the curves, which at first constantly changed their forms and were plentifully furnished with peaks and valleys, will gradually become more like each other, as also simpler and more smooth, so that at last, when we have a very large but finite number of observations, we cannot distinguish the successive figures we have drawn from one another. We may thus directly construct curves of errors, which may be approved as pictures of curves of presumptive errors, but in order to do so millions of repetitions, rather than thousands, are certainly required.

If from curves of actual errors for small numbers we are to draw conclusions as to the curve of presumptive errors, we must guess, but at the same time support our guess, partly by an estimate of how great irregularities we may expect in a curve of actual errors for the given number, partly by developing our feeling for the form of regular curves of that sort, as we must suppose that the curves of presumptive errors will be very regular. In both respects we must get some practice, but this is easy and interesting

Without feeling tied down to the particular points that determined the curve of actual errors, we shall nevertheless try to approach them, and especially not allow many large deviations on the same side to come together. We can generally regard as large deviations (the reason why will be mentioned in the chapter on the Theory of Probabilities) those that cause greater errors, as compared with the absolute frequency of the result in question, than the square root of that number (more exactly $\sqrt{h\frac{n-h}{n}}$, where h is the frequency of the result, n the number of all repetitions). But even deviations two or three times as great as this ought not always to be avoided, and we may be satisfied, if only one third of the deviations of the determining points must be called large. We may use

the word "adjustment" (graphical) to express the operation by which a curve of presumptive errors is determined. (Comp. § 64) The adjustment is called an over-adjustment, if we have approached too near to some imaginary ideal, but if we have kept too close to the curve of actual errors, then the curve is said to be under-adjusted

Our second guide, the regularity of the curve of errors, is as an æsthetical notion of a somewhat vague kind. The continuity of the curve is an essential condition, but it is not sufficient. The regularity here is of a somewhat different kind from that seen in the examples of simple, continuous curves with which students more especially become acquainted. The curves of errors get a peculiar stamp, because we would never select the essential circumstances of the observation so absurdly that the deviations could become indefinitely large. Nor would we without necessity retain a form of observation which might bring about discontinuity. It follows that to the abscissae which indicate very large deviations, must correspond rapidly decreasing ordinates The curve of errors must have the axis of abscissae as an asymptote, both to the right and the left. All frequency being positive, where the curve of errors deviates from the axis of abscissae, it must exclusively keep on the positive side of the latter. It must therefore more or less get the appearance of a bow, with the axis of abscissae for the string In order to train the eye for the apprehension of this sort of regularity, we recommend the study of figs. 2 & 3, which represent curves of errors of typical forms, exponential and binomial (comp. the next chapter, p. 16, seqq.), and a comparison of them with figures which, like Nr. 1, are drawn from actual observations without any adjustment.

The best way to acquire practice in drawing curves of errors, which is so important that no student ought to neglect it, may be to select a series of observations, for which the law of presumptive errors may be considered as known, and which is before us in tabular form.

We commence by drawing curves of actual errors for the whole series of observations; then for tolerably large groups of the same, and lastly for small groups taken at random and each containing only a few observations. On each drawing we draw also, besides the curve of actual errors, another one of the presumptive errors, on the same scale, so that the abscissae are common, and the ordinates indicate relative frequencies in proportion to the same unit of length for the total number. The proportions ought to be chosen so that the whole part of the axis of abscissae which deviates sensibly from the curve, is between 2 and 5 times as long as the largest ordinate of the curve.

Prepared by the study of the differences between the curves, we pass on at last to the construction of curves of presumptive errors immediately from the scattered points of the curve which correspond to the observed frequencies. In this construction we must not consider ourselves obliged to reproduce the curve of presumptive errors which we may

know beforehand, our task is to represent the observations as nearly as possible by means of a curve which is as smooth and regular as that curve.

The following table of 500 results, got by a game of patience, may be treated in this way as an exercise.

Result	Actual frequency for groups of																									For all 500	The law of errors of the method, interpolated	Result
	25 repetitions																				100 repetitions							
	I				II				III				IV				V				I	II	III	IV	V			
7	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	0	1	3	0·0003	7
8	0	0	0	1	0	2	2	0	0	1	1	0	0	0	0	0	0	0	0	0	1	4	2	0	0	7	0·0019 0·0071	8
9	1	3	1	1	5	3	1	1	3	2	2	0	3	1	0	3	1	1	2	1	6	10	7	7	5	35	0·0192 0·0392	9
10	9	2	9	5	6	6	6	4	5	4	8	3	3	3	5	6	4	6	3	4	25	22	20	17	17	101	0·0636 0·0859	10
11	3	6	3	3	3	6	4	4	5	5	3	5	3	7	2	5	5	6	3	8	15	17	18	17	22	89	0·1005 0,1054	11
12	8	5	3	4	3	3	2	8	3	7	4	6	5	4	6	5	3	3	5	7	20	16	20	20	18	94	0·1021 0·0934	12
13	2	4	4	3	6	3	3	1	4	1	1	3	5	4	3	6	7	3	6	1	13	13	9	18	17	70	0·0823 0·0705	13
14	1	2	2	4	1	0	2	3	2	1	2	4	3	5	4	0	4	0	2	4	9	6	9	12	10	46	0·0591 0·0485	14
15	0	1	2	2	1	1	3	2	3	2	2	3	1	1	3	0	0	2	1	0	5	7	10	5	3	30	0·0387 0·0298	15
16	1	2	0	1	0	1	1	0	0	1	0	1	2	0	0	0	0	3	2	0	4	2	2	2	5	15	0·0216 0·0145	16
17	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	1	0	0	2	1	4	0·0088 0·0046	17
18	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	1	1	2	0	1	5	0·0020 0·0007	18
19	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0·0002 0·0000	19
Total	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	100	100	100	100	100	500	0·9999	

The law of presumptive errors here given is not the direct result of free-hand construction; but the curve so got has been improved by interpolation of the logarithms of its statements of the relative frequencies, together with the formation of mean numbers for the deviations, a proceeding which very often will give good results, but which is not strictly necessary. By this we can also determine the functional law of errors (Comp. the

next chapter). The equation of the curve is.

$$\operatorname{Log} y = 2{\cdot}0228 + 0{\cdot}0030\,(x-11) - 0{\cdot}6885\,(x-11)^2 + 0{\cdot}01515\,(x-11)^3 - 0{\cdot}001675\,(x-11)^4$$

§ 15. By the study of many curves of presumptive errors, and especially such as represent ideal functional laws of errors, we cannot fail to get the impression that there exists a typical form of curves of errors, which is particularly distinguished by symmetry. Familiarity with this form is useful for the construction of curves of presumptive errors. But we must not expect to get it realised in all cases. For this reason I have considered it important to give, alongside of the typical curves, an example taken from real observations of a skew curve of errors, which in consequence of its marked want of symmetry deviates considerably from the typical form. Fig. 4 shows this last mentioned law of presumptive errors.

Deviation from the typical form does not indicate that the observations are not good But it may become so glaring that we are forced by it to this conclusion. If, for instance, between the extreme values of repetitions — abscissae — there are intervals which are as free from finite ordinates as the space beyond the extremes, so that the curve of errors is divided into two or several smaller curves of errors beside one another, there can scarcely be any doubt that we have not a series of repetitions proper, but a combination of several; that is to say, different methods of observation have been used and the results mixed up together. In such cases we cannot expect that the law of large numbers will remain in force, and we had better, therefore, reject such observations, if we cannot retain them by tracing out the essential circumstances which distinguish the groups of the series, but have been overlooked.

§ 16. When a curve of presumptive errors is drawn, we can measure the magnitude of the ordinate for any given abscissa; so far then we know the law of errors perfectly, by means of the curve of errors, but certainly in the tabular form only, with all its copiousness Whether we can advance further depends on, whether we succeed in interpolating in the table so found, and particularly on, whether we can, either from the table or direct from the curve of errors, by measurement obtain a comparatively small number of constants, by which to determine the special peculiarities of the curve.

By interpolating, by means of Newton's formula, the logarithms of the frequencies, or by drawing the curves of errors with the logarithms of the frequencies as ordinates, we often succeed, as above mentioned, in giving the curve the form of a parabola of low (and always even) degree.

Still easier is it to make use of the circumstance that fairly typical curves of errors show a single maximum ordinate, and an inflexion on each side of it, near which the curve for a short distance is almost rectilinear. By measuring the co-ordinates of the maximum point and of the points of inflexion, we shall get data sufficient to enable us to

draw a curve of errors which, as a rule, will deviate very little from the original. All this, however, holds good only of the curves of presumptive errors. With the actual ones we cannot operate in this way, and the transition from the latter to the former seems in the meantime to depend on the eye's sense of beauty.

V. FUNCTIONAL LAWS OF ERRORS,

§ 17. Laws of errors may be represented in such a way that the frequency of the results of repetitions is stated as a mathematical function of the number, or numbers, expressing the results. This method only differs from that of curves of errors in the circumstance that the curve which represents the errors has been replaced by its mathematical formula; the relationship is so close that it is difficult, when we speak of these two methods, to maintain a strict distinction between them.

In former works on the theory of observations the functional law of errors is the principal instrument. Its source is mathematical speculation; we start from the properties which are considered essential in ideally good observations. From these the formula for the typical functional law of errors is deduced; and then it remains to determine how to make computations with observations in order to obtain the most favourable or most probable results.

Such investigations have been carried through with a high degree of refinement; but it must be regretted that in this way the real state of things is constantly disregarded. The study of the curves of actual errors and the functional forms of laws of actual errors have consequently been too much neglected.

The representation of functional laws of errors, whether laws of actual errors or laws of presumptive errors founded on these, must necessarily begin with a table of the results of repetitions, and be founded on interpolation of this table. We may here be content to study the cases in which the arguments (i. e. the results of the repetitions) proceed by constant differences, and the interpolated function, which gives the frequency of the argument, is considered as the functional law of errors. Here the only difficulty we encounter is that we cannot directly employ the usual Newtonian formula of interpolation, as this supposes that the function is an integral algebraic one, and gives infinite values for infinite arguments, whether positive or negative, whereas here the frequency of these infinite arguments must be $= 0$. We must therefore employ some artifice, and an obvious one is to interpolate, not the frequency itself, y, but its reciprocal, $\frac{1}{y}$. This, however, turns out to be inapplicable; for $\frac{1}{y}$ will often become infinite for finite arguments, and will, at any rate, increase much faster than any integral function of low degree.

But, as we have already said, the interpolation generally succeeds, when we apply it to the logarithm of the frequency, assuming that

$$\operatorname{Log} y = a + bx + cx^2 + \ldots + gx^{2n},$$

where the function on the right side begins with the lowest powers of the argument x, and ends with an even power whose coefficient g must be *negative*. Without this latter condition the computed frequency,

$$y = 10^{a+bx+cx^2+\ldots+gx^{2n}}, \tag{1}$$

would again become infinitely great for $x = \pm\infty$. That the observed frequency is often $= 0$, and its logarithm $-\infty$ like $\frac{1}{y}$, does no harm. Of course we must leave out these frequencies of the interpolation, or replace them by very small finite frequencies, a few of which it may become necessary to select arbitrarily. As a rule it is possible to succeed by this means. In order to represent a given law of actual errors in this way, we must, according to the rule of interpolation, determine the coefficients $a, b, c, \ldots g$, whose number must be at least as large as that of the various results of repetitions with which we have to deal. This determination, of course, is a troublesome business.

Here also we may suppose that the law of presumptive errors is simpler than that of the actual errors. And though this, of course, does not imply that $\log y$ can be expressed by a small number of terms containing the lowest powers of x, this supposition, nevertheless, is so obvious that it must, at any rate, be tried before any other.

§ 18. Among these, the simplest case, namely that in which $\operatorname{Log} y$ is a function of x of the second degree

$$\operatorname{Log} y = a + bx - cx^2,$$

gives us the typical form for the functional law of errors, and for the curve of errors, or with other constants

$$y = he^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} = h10^{-0\cdot21715\left(\frac{x-m}{n}\right)^2}, \tag{2}$$

where

$$e = 1 + \frac{1}{1} + \frac{1}{1\cdot2} + \frac{1}{1\cdot2\cdot3} + \ldots = 2\cdot71828.$$

The function has therefore no other constants than those which may be interpreted as unit for the frequencies h, and as zero m and unit n for the observed values; the corresponding typical curve of errors has therefore in all essentials a fixed form.

The functional form of the typical law of errors has applications in mathematics which are almost as important as those of the exponential, logarithmic, and trigonometrical functions. In the theory of observations its importance is so great that, though it has been over-estimated by some writers, and though many good observations show presumptive as well as actual laws of errors that are not typical, yet every student must make himself perfectly familiar with its properties.

Expanding the index we get

$$e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} = e^{-\frac{1}{2}\left(\frac{m}{n}\right)^2}\cdot e^{\frac{xm}{n^2}}\cdot e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}, \tag{3}$$

so that the general function resolves itself into a product of three factors, the first of which is constant, the second an ordinary exponential function, while the third remains a typical functional law of errors. Long usage reduces this form to e^{-x^2}; but this form cannot be recommended In the majority of its purely mathematical applications $e^{-\pi x^2}$ is preferable, unless (as in the whole theory of observations) the factor $\frac{1}{2}$ in the index is to be preferred on account of the resulting simplification of most of the derived formulæ.

The differential coefficients of $e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}$ with regard to x are

$$\left.\begin{aligned}
De^{-\frac{1}{2}\left(\frac{x}{n}\right)^2} &= -n^{-2}xe^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}\\
D^2e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2} &= n^{-4}(x^2-n^2)e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}\\
D^3e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2} &= -n^{-6}(x^3-3n^2x)e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}\\
D^4e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2} &= n^{-8}(x^4-3n^2\cdot 2x^2+1\cdot 3n^4)e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}\\
D^5e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2} &= -n^{-10}(x^5-5n^2\cdot 2x^3+3\cdot 5n^4x)e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}\\
D^6e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2} &= n^{-12}(x^6-5n^2\cdot 3x^4+3\cdot 5n^4\cdot 3x^2-1\cdot 3\cdot 5n^6)e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}
\end{aligned}\right\} \tag{4}$$

The law of the numerical coefficients (products of odd numbers and binomial numbers) is obvious. The general expression of $D^re^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}$ can be got from a comparison of the coefficients to $(-m)^r$ of the two identical series for equation (3), one being the Taylor series, the other the product of $e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2}$ and the two exponential series with m^2 and m as arguments. It can also be induced from the differential equation

$$n^2D^{r+2}\varphi + xD^{r+1}\varphi + (r+1)D^r\varphi = 0.$$

Inversely, we obtain for the products of the typical law of errors by powers of x

$$\left.\begin{aligned}
x\varphi &= -n^2D\varphi\\
x^2\varphi &= n^4D^2\varphi+n^2\varphi\\
x^3\varphi &= -n^6D^3\varphi-3n^4D\varphi\\
x^4\varphi &= n^8D^4\varphi+6n^6D^2\varphi+3n^4\varphi\\
x^5\varphi &= -n^{10}D^5\varphi-10n^8D^3\varphi-15n^6D\varphi\\
x^6\varphi &= n^{12}D^6\varphi+15n^{10}D^4\varphi+45n^8D^2\varphi+15n^6\varphi\\
\varphi &= e^{-\frac{1}{2}\left(\frac{x}{n}\right)^2},
\end{aligned}\right\} \tag{5}$$

the numerical coefficients being the same as above (4). This proposition can be demonstrated by the identical equation $n^{-2}x^{r+1}\varphi = -D(x^r\varphi)+rx^{r-1}\varphi$.

By means of these formulæ every product of any integral rational function by

exponential functions and functional typical laws of errors can be reduced to the form

$$k_0\varphi - \frac{k_1}{\underline{|1}} D\varphi + \frac{k_2}{\underline{|2}} D^2\varphi - \frac{k_3}{\underline{|3}} D^3\varphi + \dots, \tag{6}$$

where

$$\varphi = e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2},$$

and thus they can easily be differentiated and integrated. Every quadrature of this form can be reduced to

$$f_1(x)e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2} + f_2(x)\int e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}dx,$$

where $f_1(x)$ and $f_2(x)$ are integral rational functions; thus a very large class of problems can be solved numerically by aid of the following table of the typical or exponential functional law of errors, $\eta = e^{-\frac{1}{2}z^2}$, together with the table of its integral $\int_0^z \eta dz$.

z	$\int_0^z \eta dz$	$\eta = e^{-\frac{1}{2}z^2}$	$\frac{d\eta}{dz}$	$\frac{d^2\eta}{dz^2}$	$\frac{d^3\eta}{dz^3}$	$\frac{d^4\eta}{dz^4}$
0·0	0·00000	1·0000	0·000	−1·00	0·0	3
0·1	0·09983	·9950	−·100	−·99	·3	3
0·2	0·19867	·9802	−·196	−·94	·6	3
0·3	0·29556	·9560	−·287	−·87	8	2
0·4	0·38968	·9231	−·369	−·78	1·0	2
0·5	0·47993	·8825	−·441	−·66	1·2	1
0·6	0·56586	·8353	−·501	−·53	1·3	1
0·7	0·64680	·7827	−·548	−·40	1·4	0
0·8	0·72227	·7261	−·581	−·26	1·4	−0
0·9	0·79194	·6670	−·600	−·13	1·3	−1
1·0	0·85562	0·6065	−0·607	0·00	1·2	−1
1·1	0·91325	·5461	−·601	·11	1·1	−2
1·2	0·96458	·4868	−·584	·21	·9	−2
1·3	1·01067	·4296	−·558	·30	·7	−2
1·4	1·05089	·3753	−·525	·36	·5	−2
1·5	1·08585	·3247	−·487	·41	·4	−2
1·6	1·11595	·2780	−·445	·43	·2	−2
1·7	1·14161	·2357	−·401	·45	·0	−1
1·8	1·16325	·1979	−·356	·44	−·1	−1
1·9	1·18139	·1645	−·313	·43	−·2	−1
2·0	1·19629	0·1353	−0·271	0·41	−·3	−1
2·1	1·20858	·1103	−·232	·38	−·3	−0
2·2	1·21846	·0889	−·196	·34	−·4	−0
2·3	1·22643	·0710	−·163	·30	−·4	−0
2·4	1·23277	0·0561	−0·135	0·27	−0·4	0
2·5	1·23775	·0439	−·110	·23	−·4	0
2·6	1·24163	·0340	−·089	·20	−·3	0
2·7	1·24462	·0261	−·071	·16	−·3	0
2·8	1·24691	·0198	−·056	·14	−·3	0
2·9	1·24864	·0149	−·043	·11	−·2	0
3·0	1·24993	0·0111	−0·033	0·09	−0·2	0·3
3·1	1·25089	·0082	−·025	·07	−·2	0
3·2	1·25159	·0060	−·019	·06	−·1	0
3·3	1·25210	·0043	−·014	·04	−·1	0
3·4	1·25247	·0031	−·011	·03	−·1	0
3·5	1·25273	·0022	−·008	·02	−·1	0
3·6	1·25292	·0015	−·006	·02	−·1	0
3·7	1·25304	·0011	−·004	·01	−·0	0
3·8	1·25313	·0007	−·003	·01	−·0	0
3·9	1·25319	·0005	−·002	·01	−·0	0
4·0	1·25323	0·0003	−0·001	0·01	$-\frac{1}{15}$	$\frac{1}{11}$
4·1	1·25326	·0002	−·001	·00		
4·2	1·25328	·0001	−·001	·00		
4·3	1·25329	·0001	−·000	·00		
4·4	1·25330	·0001	−·000	·00		
4·5	1·25331	·0000	−·000	·00		
∞	$\sqrt{\frac{1}{2}\pi}$	·0000	−·000	·00	0	0

Here η, $\frac{d^2\eta}{dz^2}$, $\frac{d^4\eta}{dz^4}$ are, each of them, the same for positive and negative values of z; the other columns of the table change signs with z.

The interpolations are easily worked out by means of Taylor's theorem:

$$\eta_{(z+\zeta)} = \eta + \frac{d\eta}{dz}\cdot\zeta + \tfrac{1}{2}\frac{d^2\eta}{dz^2}\cdot\zeta^2 + \tfrac{1}{6}\frac{d^3\eta}{dz^3}\cdot\zeta^3 + \tfrac{1}{24}\frac{d^4\eta}{dz^4}\cdot\zeta^4 + \dots \tag{7}$$

and

$$\int_0^{z+\zeta}\eta dz = \int_0^z \eta dz + \eta\cdot\zeta + \tfrac{1}{2}\frac{d\eta}{dz}\cdot\zeta^2 + \tfrac{1}{6}\frac{d^2\eta}{dz^2}\cdot\zeta^3 + \tfrac{1}{24}\frac{d^3\eta}{dz^3}\cdot\zeta^4 + \tfrac{1}{120}\frac{d^4\eta}{dz^4}\cdot\zeta^5 + \dots \tag{8}$$

The typical form for the functional law of errors (2) shows that the frequency is always positive, and that it arranges itself symmetrically about the value $x = m$, for which the frequency has its maximum value $y = h$. For $x = m \pm n$ the frequency is $y = h\cdot 0{\cdot}60653$. The corresponding points in the curve of errors are the points of inflexion. The area between the curve of errors and the axis of abscissae, reckoned from the middle to $x = m \pm n$, will be $nh\cdot 0{\cdot}85562$; and as the whole area from one asymptote to the other is $nh\sqrt{2\pi} = nh\cdot 2{\cdot}50663$, only $nh\cdot 0{\cdot}39769$ of it falls outside either of the inflexions, consequently not quite that sixth part (more exactly 16 per ct.) which is the foundation of the rule, given in § 11, as to the limit between the great and small errors.

The above table shows how rapidly the function of the typical law of errors decreases toward zero. In almost all practical applications of the theory of observations $e^{-\frac{1}{2}z^2} = 0$, if only $z > 5$. Theoretically this superior assymptotical character of the function is expressed in the important theorem that, for $z = \pm\infty$, not only $e^{-\frac{1}{2}z^2}$ itself is $= 0$ but also all its differential coefficients; and that, furthermore, all products of this function by every algebraic integral function and by every exponential function, and all the differential quotients of these products, are equal to zero.

In consequence of this theorem, the integral $\int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2}dz = \sqrt{2\pi}$ can be computed as the sum of equidistant values of $e^{-\frac{1}{2}z^2}$ multiplied by the interval of the arguments without any correction. This simple method of computation is not quite correct, the underlying series for conversion of a sum into an integral being only semiconvergent in this case; for very large intervals the error can be easily stated, but as far as intervals of one unit the numbers taken out of our table are not sufficient to show this error.

If the curve of errors is to give relative frequency directly, the total area must be $1 = nh\sqrt{2\pi}$; h consequently ought to be put $= \frac{1}{n\sqrt{2\pi}}$.

Problem 1. Prove that every product of typical laws of errors in the functional form $= he^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2}$, with the same independent variable x is, itself a typical law of errors. How do the constants h, m, and n change in such a multiplication?

Problem 2. How small are the frequencies of errors exceeding 2, 3, or 4 times the mean error, on the supposition of the typical law of errors?

Problem 3. To find the values of the definite integrals

$$s_r = \int_{-\infty}^{\infty} x^r e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} dx.$$

Answer: $s_{2i+1} = 0$ and $s_{2i} = 1\cdot3\cdot5\ldots(2i-1)\, n^{2i+1}\sqrt{2\pi}$.

§ 19. Nearly related to the typical or exponential law of errors in functional form are the binomial functions, which are known from the coefficients of the terms of the nth power of a binomial, regarded as a function of the number x of the term.

n	$x =$ 0	1	2	3	4	5	6	7
0	1							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	
7	1	7	21	35	35	21	7	1
8	1	8	28	56	70	56	28	8
9	1	9	36	84	126	126	84	36
10	1	10	45	120	210	252	210	120
11	1	11	55	165	330	462	462	330
12	1	12	66	220	495	792	924	792
13	1	13	78	286	715	1287	1716	1716
14	1	14	91	364	1001	2002	3003	3432

For integral values of the argument the binomial function can be computed directly by the formula

$$\left.\begin{aligned} \beta_n(x) &= \frac{1\cdot2\cdot3\ldots n}{1\cdot2\cdot3\ldots x\cdot1\cdot2\cdot3\ldots(n-x)} = \beta_n(n-x) \\ &= \frac{n(n-1)\ldots(n-x+1)}{1\cdot2\ldots x}, \end{aligned}\right\} \tag{9}$$

When the binomial numbers for n are known, those for $n+1$ are easily found by the formula

$$\beta_{n+1}(x) = \beta_n(x) + \beta_n(x-1). \tag{10}$$

By substitution according to (9) we easily demonstrate the proposition that, for

any integral values of n, r, and t

$$\beta_n(t)\beta_{r-t}(r) = \beta_n(r)\cdot\beta_{n-r}(t), \tag{11}$$

which means that, when the trinomial $(a+b+c)^n$ is developed, it is indifferent whether we consider it to be $((a+b)+c)^n$ or $(a+(b+c))^n$.

For fractional values of the argument x, the binomial function $\beta_n(x)$ can be taken in an infinity of different ways, for instance by

$$\beta_0(x) = \frac{\sin \pi x}{\pi x}.$$

This formula results from a direct application of Lagrange's method of interpolation, and leads by (10) to the more general formula

$$\beta_n(x) = \frac{1\cdot 2\ldots n}{(1-x)(2-x)\ldots(n-x)}\,\frac{\sin \pi x}{\pi x}. \tag{12}$$

This species of binomial function may be considered the simplest possible, and has some importance in pure mathematics; but as an expression of frequencies of observed values, or as a law of errors, it is inadmissible because, for $x > n$ or x negative, it gives negative values alternating with positive values periodically.

This, however, may be remedied. As $\beta_0(x)$ has no other values than 0 and 1, when x is integral, we can put for instance

$$\left.\begin{aligned}
&\beta_0(x) = \left(\frac{\sin \pi x}{\pi x}\right)^2; \\
&\text{by (10) then} \\
&\beta_1(x) = \left(\frac{1}{x^2}+\frac{1}{(x-1)^2}\right)\frac{\sin^2 \pi x}{\pi^2} \\
&\beta_2(x) = \left(\frac{1}{x^2}+\frac{2}{(x-1)^2}+\frac{1}{(x-2)^2}\right)\frac{\sin^2 \pi x}{\pi^2}
\end{aligned}\right\} \tag{13}$$

Here the values of the binomial function are constantly positive or 0. But this form is cumbersome, and although for $x = \infty$ the function and its principal coefficients are $= 0$, this property is lost here, when we multiply by integral algebraic or by exponential functions.

These unfavourable circumstances detract greatly from the merits of the binomial functions as expressions for continuous laws of errors.

When, on the contrary, the observations correspond only to integral values of the argument, the original binomial functions are most valuable means for treating them. That $\beta_n(x) = 0$, if $x > n$ or negative, is then of great importance. But this case must be referred to special investigations.

§ 20. To represent non-typical laws of errors in functional form we have now the choice between at least three different plans:

1) the formula (1) or

$$y = e^{\alpha+\beta x+\gamma x^2+\cdots+\kappa x^r},$$

2) the products of integral algebraic functions by a typical function or (6)

$$y = k_0\varphi - \frac{k_1}{\underline{|1}} D\varphi + \frac{k_2}{\underline{|2}} D^2\varphi - \frac{k_3}{\underline{|3}} D^3\varphi + \cdots, \qquad \varphi = e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2},$$

3) a sum of several typical functions

$$y = \Sigma_1^r h_i e^{-\frac{1}{2}\left(\frac{x-m_i}{n_i}\right)^2}. \tag{14}$$

This account of the more prominent among the functional forms, which we have at our disposal for the representation of laws of errors, may prove that we certainly possess good instruments, by means of which we can even in more than one form find general series adapted for the representation of laws of errors. We do not want forms for the series, required in theoretical speculations upon laws of errors; nor is the exact representation of the actual frequencies more than reasonably difficult. If anything, we have too many forms and too few means of estimating their value correctly.

As to the important transition from laws of actual errors to those of presumptive errors, the functional form of the law leaves us quite uncertain. The convergency of the series is too irregular, and cannot in the least be foreseen.

We ask in vain for a fixed rule, by which we can select the most important and trustworthy forms with limited numbers of constants, to be used in predictions. And even if we should have decided to use only the typical form by the laws of presumptive errors, we still lack a method by which we can compute its constants. The answer, that the "adjustment" of the law of errors must be made by the "method of least squares", may not be given till we have attained a satisfactory proof of that method; and the attempts that have been made to deduce it by speculations on the functional laws of errors must, I think, all be regarded as failures.

VI. LAWS OF ERRORS

EXPRESSED BY SYMMETRICAL FUNCTIONS.

§ 21. All constants in a functional law of errors, every general property of a curve of errors or, generally, of a law of numerical errors, must be symmetrical functions of the several results of the repetitions, i. e. functions which are not altered by interchanging two or more of the results. For, as all the values found by the repetitions correspond to the same essential circumstances, no interchanging whatever can have any influence on the law of errors. Conversely, any symmetrical function of the values of the

observations will represent some property or other of the law of errors. And we must be able to express the whole law of errors itself by every such collection of symmetrical functions, by which every property of the law of errors can be expressed as unambiguously as by the very values found by the repetitions.

We have such a collection in the coefficients of that equation of the n^{th} degree, whose roots are the n observed values. For if we know these coefficients, and solve the equation, we get an unambiguous determination of all the values resulting from the repetitions, i. e. the law of errors But other collections also fulfil the same requirements; the essential thing is that the n symmetrical functions are rational and integral, and that one of them has each of the degrees $1, 2 \dots n$, and that none of them can be deduced from the others.

The collection of this sort that is easiest to compute, is *the sums of the powers*. With the observed values

$$o_1, o_2, o_3, \dots o_n$$

we have

$$\left.\begin{aligned} s_0 &= o_1^0 + o_2^0 + \dots + o_n^0 = n \\ s_1 &= o_1 + o_2 + \dots + o_n \\ s_2 &= o_1^2 + o_2^2 + \dots + o_n^2 \\ &\dots\dots\dots \\ s_r &= o_1^r + o_2^r + \dots + o_n^r \end{aligned}\right\} \qquad (15)$$

and the fractions $\frac{s_r}{n}$ may also be employed as an expression for the law of errors; it is only important to reduce the observations to a suitable zero which must be an average value of $o_1 \dots, o_n$; for if the differences between the observations are small, as compared with their differences from the average, then

$$\frac{s_1}{s_0}, \sqrt{\frac{s_2}{s_0}}, \dots \sqrt[r]{\frac{s_r}{s_0}}$$

may become practically identical, and therefore unable to express more than one property of the law of errors.

From a well known theorem of the theory of symmetrical functions, the equations

$$\begin{aligned} 1 + a_1\omega + a_2\omega^2 + \dots &= (1 - o_1\omega)(1 - o_2\omega)\dots(1 - o_n\omega) \\ &= e^{\Sigma \log(1 - o_r\omega)} \\ &= e^{-(s_1\omega + \frac{1}{2}s_2\omega^2 + \frac{1}{3}s_3\omega^3 + \dots)}, \end{aligned}$$

which are identical with regard to every value of ω, we learn that the sum of the powers s_r can be computed without ambiguity, if we know the coefficients a_r of the equation, whose roots are the n observations; and vice versâ, by differentiating the last equation

with regard to ω, and equating the coefficients we get

$$\left.\begin{array}{l} 0 = a_1 + s_1 \\ 0 = 2a_2 + a_1 s_1 + s_2 \\ \dots\dots\dots\dots\dots\dots \\ 0 = n a_n + a_{n-1} s_1 + \dots + a_1 s_{n-1} + s_n \end{array}\right\} \qquad (16)$$

from which the coefficients a_n are unambiguously and very easily computed, when the s_r are directly calculated

§ 22. But from the sums of powers we can easily compute also another serviceable collection of symmetrical functions, which for brevity we shall call *the half-invariants.*

Starting from the sums of powers s_r, these can be defined as μ_1, μ_2, μ_3, by the equation

$$s_0 e^{\frac{\mu_1}{\underline{|1}}\tau + \frac{\mu_2}{\underline{|2}}\tau^2 + \frac{\mu_3}{\underline{|3}}\tau^3 \dots} = s_0 + \frac{s_1}{\underline{|1}}\tau + \frac{s_2}{\underline{|2}}\tau^2 + \frac{s_3}{\underline{|3}}\tau^3 + \dots, \qquad (17)$$

which we suppose identical with regard to τ.

As $s_r = \Sigma o^r$, this can be written

$$s_0 e^{\frac{\mu_1}{\underline{|1}}\tau + \frac{\mu_2}{\underline{|2}}\tau^2 + \frac{\mu_3}{\underline{|3}}\tau^3 + \dots} = e^{o_1\tau} + e^{o_2\tau} + \dots e^{o_n\tau}. \qquad (18)$$

By developing the first term of (17) as $\Sigma k_r \tau_r$ and equating the coefficients of each power of τ, we get each $\frac{s_r}{s_0}$ expressed as a function of $\mu_1 \dots \mu_r$:

$$\left.\begin{array}{l} s_1 = s_0 \mu_1 \\ s_2 = s_0 (\mu_2 + \mu_1^2) \\ s_3 = s_0 (\mu_3 + 3\mu_2\mu_1 + \mu_1^3) \\ s_4 = s_0 (\mu_4 + 4\mu_3\mu_1 + 3\mu_2^2 + 6\mu_2\mu_1^2 + \mu_1^4) \\ \dots\ \dots \end{array}\right\} \qquad (19)$$

Taking the logarithms of (17) we get

$$\frac{\mu_1}{\underline{|1}}\tau + \frac{\mu_2}{\underline{|2}}\tau^2 + \frac{\mu_3}{\underline{|3}}\tau^3 + \dots = \log\left(1 + \frac{s_1}{s_0}\frac{\tau}{\underline{|1}} + \frac{s_2}{s_0}\frac{\tau^2}{\underline{|2}} + \frac{s_3}{s_0}\frac{\tau^3}{\underline{|3}} + \dots\right) \qquad (20)$$

and hence

$$\left.\begin{array}{l} \mu_1 = s_1 : s_0 \\ \mu_2 = (s_2 s_0 - s_1^2) : s_0^2 \\ \mu_3 = (s_3 s_0^2 - 3 s_2 s_1 s_0 + 2 s_1^3) : s_0^3 \\ \mu_4 = (s_4 s_0^3 - 4 s_3 s_1 s_0^2 - 3 s_2^2 s_0^2 + 12 s_2 s_1^2 s_0 - 6 s_1^4) : s_0^4 \\ \dots\dots \end{array}\right\} \qquad (21)$$

The general law of the relation between the μ and s is more easily understood through the equations

$$
\left.\begin{aligned}
s_1 &= \mu_1 s_0 \\
s_2 &= \mu_1 s_1 + \mu_2 s_0 \\
s_3 &= \mu_1 s_2 + 2\mu_2 s_1 + \mu_3 s_0 \\
s_4 &= \mu_1 s_3 + 3\mu_2 s_2 + 3\mu_3 s_1 + \mu_4 s_0 \\
&\ldots\ldots
\end{aligned}\right\} \quad (22)
$$

where the numerical coefficients are those of the binomial theorem. These equations can be demonstrated by differentiation of (17) with regard to τ, the resulting equation

$$
s_1 + \frac{s_2}{\underline{|1}}\tau + \frac{s_3}{\underline{|2}}\tau^2 + \frac{s_4}{\underline{|3}}\tau^3 + \ldots = \left(\mu_1 + \frac{\mu_2}{\underline{|1}}\tau + \frac{\mu_3}{\underline{|2}}\tau^2 + \ldots\right)\left(s_0 + \frac{s_1}{\underline{|1}}\tau + \frac{s_2}{\underline{|2}}\tau^2 + \ldots\right) \quad (23)
$$

being satisfied for all values of τ by (22).

These half-invariants possess several remarkable properties. From (18) we get

$$
s_0 e^{\frac{\mu_2}{\underline{|2}}\tau^2 + \frac{\mu_3}{\underline{|3}}\tau^3 + \ldots} = e^{(o_1 - \mu_1)\tau} + \ldots + e^{(o_n - \mu_1)\tau} \quad (24)
$$

consequently any transformation $o' = o + c$, any change of the zero of all observations $o_1 \ldots o_n$, affects only μ_1 in the same manner, but leaves $\mu_2, \mu_3, \mu_4, \ldots$ unaltered; any change of the unit of all observations can be compensated by the reciprocal change of the unit of τ, and becomes therefore indifferent to $\mu_2\tau^2, \mu_3\tau^3, \ldots$

Not only the ratios

$$
\frac{s_1}{s_0}, \; \frac{s_2}{s_0}, \ldots \frac{s_n}{s_0}
$$

but also the half-invariants have the property which is so important in a law of errors, of remaining unchanged when the whole series of repetitions is repeated unchanged.

We have seen that the typical character of a law of errors reveals itself in the elegant functional form

$$
\varphi(x) = e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2}.
$$

Now we shall see that it is fully as easy to recognize the typical laws of errors by means of their half-invariants. Here the criterion is that $\mu_r = 0$ if $r \geq 3$, while $\mu_1 = m$ and $\mu_2 = n^2$. This remarkable proposition has originally led me to prefer the half-invariants to every other system of symmetrical functions; it is easily demonstrated by means of (5), if we take m for the zero of the observations.

We begin by forming the sums of powers s_r of that law of errors where the frequency of an observed x is proportional to $\varphi(x) = e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2}$; as this law is continuous we get

$$
s_r = \int_{-\infty}^{+\infty} x^r \varphi(x)\, dx.
$$

For every differential coefficient $D^m\varphi(x)$ we have

$$\int_{-\infty}^{+\infty} D^m\varphi(x)\cdot dx = D^{m-1}\varphi(\infty) - D^{m-1}\varphi(-\infty) = 0,$$

consequently we learn from (5) that $s_{2r+1} = 0$, but

$$\begin{aligned} s_2 &= 1\cdot n^2 s_0 \\ s_4 &= 1\cdot 3\cdot n^4 s_0 \\ s_6 &= 1\cdot 3\cdot 5\cdot n^6 s_0 \\ &\ldots\ldots \end{aligned}$$

(compare problem 3, § 18). Now the half-invariants can be found by (22) or by (17). If we use (22) we remark that $s_{2r} = n^2(2r-1)s_{2r-2}$; then writing for (22)

$$\begin{aligned}
s_1 &= \mu_1 s_0 &&= 0\\
s_2 - \mu_2 s_0 &= \mu_1 s_1 &&= 0\\
s_3 - 2\mu_2 s_1 &= \mu_1 s_2 + \mu_3 s_0 &&= 0\\
s_4 - 3\mu_2 s_2 &= \mu_1 s_3 + 3\mu_3 s_1 + \mu_4 s_0 &&= 0\\
s_5 - 4\mu_2 s_3 &= \mu_1 s_4 + 6\mu_3 s_2 + 4\mu_4 s_1 + \mu_5 s_0 &&= 0\\
s_6 - 5\mu_2 s_4 &= \mu_1 s_5 + 10\mu_3 s_3 + 10\mu_4 s_2 + 5\mu_5 s_1 + \mu_6 s_0 &&= 0
\end{aligned}$$

we see that the solution is $\mu_2 = n^2$ and $\mu_1 = \mu_3 = \mu_4 = \ldots = 0$.

By (17) we get

$$e^{\frac{\mu_1}{|\underline{1}}\tau + \frac{\mu_2}{|\underline{2}}\tau^2 + \frac{\mu_3}{|\underline{3}}\tau^3 + \ldots} = 1 + \frac{(n\tau)^2}{2\,|\underline{1}} + \frac{(n\tau)^4}{4\,|\underline{2}} + \ldots$$

$$= e^{\frac{n^2\tau^2}{2}}.$$

Equating the coefficients of τ^r we get here also $\mu_1 = 0 = m$, $\mu_2 = n^2$, $\mu_r = 0$ if $r \geqq 3$.

If we wish to demonstrate this important proposition without change of the zero, and without the use of the equations (3) whose general demonstration is somewhat difficult, we can commence by the lemma that, for each integral and positive value of r, and also for $r = 0$, we have for the typical law of errors

$$s_{r+1} = m s_r + r n^2 s_{r-1}.$$

The function $\psi(x) = n^2 x^r e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2}$ is equal to zero both for $x = \infty$ and for $x = -\infty$; if we now between these limits integrate its differential equation

$$\frac{d\psi(x)}{dx} = \left(r n^2 x^{r-1} - (x-m)x^r\right)e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2},$$

we get

$$0 = -s_{r+1} + m s_r + r n^2 s_{r-1},$$

where

$$s_r = \int_{-\infty}^{+\infty} x^r e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} dx.$$

If we now from (22) subtract, term by term, the equations

$$\begin{aligned} s_1 &= ms_0 \\ s_2 &= ms_1 + n^2 s_0 \\ s_3 &= ms_2 + 2n^2 s_1 \\ s_4 &= ms_3 + 3n^2 s_2 \\ &\cdot \;\cdot \;\cdot \;\cdot \end{aligned}$$

it is obvious that $\mu_1 - m = 0$, $\mu_2 = n^2$, $\mu_3 = \mu_4 = \ldots = 0$

By computation of μ_1 and μ_2 we find consequently, in the simplest way, the constants of a typical law of errors.

If the law of errors deviates only a little from the typical form, μ_3, μ_4, etc., will also, all of them, be relatively small numbers; and each of them may be either positive or negative.

On the whole, a law of errors can be determined without ambiguity by the values $\mu_1, \mu_2, \ldots \mu_r$, r being the number of repetitions. From any such μ's we can compute the sums of the powers s unambiguously, and from these again the coefficients of the equation whose roots are the observed values.

But for real laws of errors it is a necessary condition that no imaginary root can be admitted. If an infinite number of repetitions is considered, the equation ceases to be algebraic, and then the convergency of the series necessary for its solution is a further condition.

§ 23. *The mean value* $\mu_1 = \frac{s_1}{s_0} = \frac{o_1 + o_2 + \ldots + o_n}{n}$ is always greater than the least, less than the greatest of the observed values $o_1, o_2, \ldots, o_n$; under typical circumstances we shall find almost the same number of greater and less values of the observations. The majority of them lie rather near to μ_1; only few very distant from it. The mean value is the *simplest* representative of what is common in a series of values found by repetition; its application as such is most likely exceedingly old, and marks in the history of science the first trace of a theory of observations.

The mean deviation, whose square is $= \mu_2$, measures the magnitude of the deviations, the uncertainty of the repeated actual observations. The square of the mean deviation is the mean of the squares of the deviations of the several observations from their mean value. By addition of

$$(o_1-\mu_1)^2 = o_1^2-2o_1\mu_1+\mu_1^2$$
$$(o_2-\mu_1)^2 = o_2^2-2o_2\mu_1+\mu_1^2$$
$$\cdots$$
$$(o_n-\mu_1)^2 = o_n^2-2o_n\mu_1+\mu_1^2$$

we get

$$\Sigma(o-\mu_1)^2 = s_2-2s_1\mu_1+s_0\mu_1^2,$$

and as $\mu_1 = \frac{s_1}{s_0}$

$$\frac{\Sigma(o-\mu_1)^2}{s_0} = \frac{s_0s_2-s_1^2}{s_0^2} = \mu_2 \qquad (25)$$

The computation of μ_2 by this formula will often be easier than by the equation (21), because s_2 in the latter must frequently be computed with more figures. There is however a middle course, which is often to be preferred to either of these methods of computation. As a *change in the zero of the observations* involves the same increase of every o and of μ_1, it will, according to (24), have no influence at all on μ_2. We select therefore as zero a convenient, round number, c, very near μ_1, and by reference to this zero the observed values are transformed to

$$o'_1 = o_1-c,\quad o'_2 = o_2-c,\quad \ldots\quad o'_n = o_n-c.$$

When s'_1 and s'_2 indicate the sums of the transformed observations, and $\mu'_1 = \mu_1-c$, then we have $\mu_1 = c+\frac{\Sigma(o-c)}{n}$ and

$$\left.\begin{aligned}\mu_2 &= \frac{s'_2}{n}-\left(\frac{s'_1}{n}\right)^2\\ &= \frac{\Sigma(o-c)^2}{n}-(\mu_1-c)^2.\end{aligned}\right\} \qquad (26)$$

We have still to mention a theorem concerning the mean deviation, which, though not useful for computation, is useful for the comprehension and further development of the idea. The square of the mean deviation μ_2 is equal to the sum of squares of the difference between each observed value and each of the others, divided by twice the square of the number. The said squares are,

$$(o_1-o_1)^2,\ (o_2-o_1)^2,\ \ldots\ (o_n-o_1)^2,$$
$$(o_1-o_2)^2,\ (o_2-o_2)^2,\ \ldots\ (o_n-o_2)^2,$$
$$\cdots$$
$$(o_1-o_n)^2,\ (o_2-o_n)^2,\ \ldots\ (o_n-o_n)^2;$$

developing each of these by the formula $(o_m-o_n)^2 = o_m^2-2o_mo_n+o_n^2$, and first adding each column separately, we find the sums

$$s_0 o_1^2 - 2 s_1 o_1 + s_2$$
$$s_0 o_2^2 - 2 s_1 o_2 + s_2$$
$$\cdot$$
$$s_0 o_n^2 - 2 s_1 o_n + s_2$$

and the sum of these

$$s_0 s_2 - 2 s_1 s_1 + s_2 s_0 = 2 (s_0 s_2 - s_1^2),$$

consequently,

$$\Sigma\Sigma (o_r - o_s)^2 = 2 s_0^2 \mu_2 \tag{27}$$

The mean deviation is greater than the least, less than the greatest of the deviations of the values of repetitions from the mean number, and less than $\sqrt{\frac{1}{2}}$ of the greatest deviation between two observed values

As to the higher half-invariants it may here be enough to state that they indicate various sorts of deviations from the typical form Skew curves of errors are indicated by the μ_{2r+1} being different from zero, peaked or flattened (divided) forms respectively by positive or negative values of μ_{4r}, and inversely by μ_{4r+2}

For these higher half-invariants we shall propose no special names But we have already introduced double names "relative frequency" and "probability" in order to accentuate the distinction between the laws of actual errors and those of presumptive errors, and the same we ought to do for the half-invariants. In what follows we shall indicate the half invariants in laws of presumptive errors by the signs λ_r instead of μ_r which will be reserved for laws of actual errors, particularly when we shall treat of the transition from laws of actual errors to those of presumptive ones For special reasons, to be explained later on, the name mean value can be used without confusion both for μ_1 and λ_1, for actual as well as for presumptive means; but instead of "mean deviation" we say "mean error", when we speak of laws of presumptive errors. Thus, if $n = s_0$,

$$\lambda_2 = \operatorname{Lim}_{n=\infty} (\mu_2)$$

is called the square of the mean error

In speculations upon ideal laws of errors, when the laws are supposed to be continuous or to relate to infinite numbers of observations, this distinction is of course insignificant

Examples

1 Professor Jul Thomsen found for the constant of a calorimeter, in experiments with pure water, in seven repetitions, the values

2049, 2047, 2045, 2053, 2053, 2046, 2049

If we take here 2050 as zero, we read the observations as

$$-1,\ -3,\ -5,\ +3,\ +3,\ -4,\ -1$$

so that

$$s'_0 = 7,\ s'_1 = -8,\ \text{and}\ s'_2 = 70;$$

consequently

$$\mu_1 = 2650 - \tfrac{8}{7} = 2649$$

$$\mu_2 = \tfrac{70}{7} - \left(-\tfrac{8}{7}\right)^2 = 9.$$

The mean deviation is consequently ± 3

2. In an alternative experiment the result is either "yes", which counts 1, or "no", which counts 0. Out of $m + n$ repetitions the m have given "yes", the n "no". What then is the expression for the law of errors in half-invariants?

$$\text{Answer}\quad \mu_1 = \frac{m}{m+n},\ \mu_2 = \frac{mn}{(m+n)^2},\ \mu_3 = \frac{mn(n-m)}{(m+n)^3},\ \mu_4 = \frac{mn(m^2-4mn+n^2)}{(m+n)^4}.$$

3. Determine the law of errors, in half-invariants, of a voting in which a voters have voted for a motion $(+1)$, c against (-1), while b have not voted (0), and examine what values for a, b, and c give the nearest approximation to the typical form.

$$\mu_1 = \frac{a-c}{a+b+c},\ \mu_2 = \frac{ab+4ca+bc}{(a+b+c)^2},\ \mu_3 = \frac{(c-a)(ab+8ca+bc-b^2)}{(a+b+c)^3},$$

$$\mu_4 = -\frac{((a+c)(a+b+c)-4(a-c)^2)(a+b+c)(2a-b+2c)+6(a-c)^4}{(a+b+c)^4}.$$

Disregarding the case when the vote is unanimous, the double condition $\mu_3 = \mu_4 = 0$ is only satisfied when one sixth of the votes is for, another sixth against, while two thirds do not give their votes. If μ_3 is to be $= 0$, without a being $= c$, $b^2 - b(a+c) - 8ac$ must be $= 0$. But then $\mu_4 = -2\mu_2\left(\frac{a-c}{a+b+c}\right)^2$, which does not disappear unless two of the numbers a, b, and c, and consequently μ_2, are $= 0$

4. Six repetitions give the quite symmetrical and almost typical law of errors, $\mu_1 = 0$, $\mu_2 = \frac{1}{3}$, $\mu_3 = \mu_4 = \mu_5 = 0$, but $\mu_6 = -\frac{2}{9}$. What are the observed values?

Answer $-1, 0, 0, 0, 0, +1$

VII. RELATIONS BETWEEN FUNCTIONAL LAWS OF ERRORS AND HALF-INVARIANTS.

§ 24. The multiplicity of forms of the laws of errors makes it impossible to write a Theory of Observations in a short manner. For though these forms are of very different value, none of them can be considered as absolutely superior to the others. The functional form which has been universally employed hitherto, and by the most prominent writers, has in my opinion proved insufficient. I shall here endeavour to replace it by the half-invariants.

But even if I should succeed in this endeavour, I am sure that not only the functional laws of errors, but even the curves of errors and the tables of frequency are too important and natural to be put completely aside without detriment

Moreover, in proposing a new plan for this theory, I have felt it my duty to explain as precisely and completely as possible its relation to the old and commonly known methods. I therefore consider it a matter of great importance that even the half-invariants, in their very definition, present a natural transition to the frequencies and to the functional law of errors

If in the equation (18)

$$n e^{\frac{\mu_1}{|1} r + \frac{\mu_2}{|2} r^2 + \cdots} = e^{o_1 r} + \cdots + e^{o_n r}$$

some of the o_i's are exactly repeated, it is of course understood that the term $e^{o_i r}$ must be counted not once but as often as o_i is repeated. Consequently, this definition of the half-invariants may, without any change of sense, be written

$$\Sigma\varphi(o_i) \, e^{\frac{\mu_1}{|1} r + \frac{\mu_2}{|2} r^2 + \frac{\mu_3}{|3} r^3 + \cdots} = \Sigma\varphi(o_i) \, e^{o_i r} \tag{28}$$

where the frequencies $\varphi(o_i)$ are given in the form of the functional law of errors For continuous laws of errors the definition must be written

$$e^{\frac{\lambda_1}{|1} r + \frac{\lambda_2}{|2} r^2 + \frac{\lambda_3}{|3} r^3 + \cdots} \int_{-\infty}^{+\infty} \varphi(o) \, do = \int_{-\infty}^{+\infty} \varphi(o) e^{or} \, do \tag{29}$$

Thus, if we know the functional law of errors and if we can perform the integrations, the half invariants may be found If, inversely, we know the λ_i, then it may be possible also to determine the functional law of errors $\varphi(o)$

Example 1 Let $\varphi(o)$ be a sum of typical functional laws of errors,

$$\varphi(o) = \Sigma h_i e^{-\frac{1}{2}\left(\frac{o - m_i}{n_i}\right)^2},$$

then $\int_{-\infty}^{+\infty} \varphi(o) \, do = \sqrt{2\pi} \, \Sigma n_i h_i$ and

$$\int_{-\infty}^{+\infty} \varphi(o) e^{or} \, do = \Sigma h_i \int_{-\infty}^{+\infty} e^{-\frac{1}{2 n_i^2}\left((o - m_i)^2 - 2 n_i^2 r o\right)} do$$

$$= \Sigma h_i e^{m_i r + \frac{n_i^2}{2} r^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{o - m_i - n_i^2 r}{n_i}\right)^2} do,$$

and consequently

$$e^{\frac{\lambda_1}{|1} r + \frac{\lambda_2}{|2} r^2 + \frac{\lambda_3}{|3} r^3 + \cdots} = \frac{\Sigma h_i n_i e^{m_i r + \frac{n_i^2}{2} r^2}}{\Sigma h_i n_i}.$$

By aid of the formulae (19) that express $\frac{s_r}{s_0}$ as functions of the λ (or μ) it is not difficult to compute the principal half-invariants. The inverse problem, to compute the m_i, n_i and h_i by means of given half-invariants is very difficult, as it results in equations of a high degree, even if only a sum of two typical functional laws of errors is in question.

Example 2. What ~~are the~~ half-invariants of a pure binomial law of errors? The observation r being repeated $\beta_n(r)$ times, we write

$$2^n e^{\frac{\mu_1}{1}\tau+\frac{\mu_2}{2!}\tau^2+\frac{\mu_3}{3!}\tau^3+\cdots} = \beta_n(0)+\beta_n(1)e^{\tau}+\cdots+\beta_n(n)e^{n\tau} = (1+e^{\tau})^n,$$

consequently

$$\left(\mu_1-\frac{n}{2}\right)\tau+\frac{\mu_2}{2!}\tau^2+\frac{\mu_3}{3!}\tau^3+\cdots = n\log\cos\frac{\tau\sqrt{-1}}{2}$$

Here the right hand side of the equation can be developed by the aid of Bernoullian numbers into a series containing only the even powers of τ, consequently

$$\mu_1 = -\frac{n}{2} \text{ and } \mu_{2r+1} = 0, \quad (r>0)$$

further

$$\mu_2 = \frac{n}{4},\; \mu_4 = -\frac{n}{8},\; \mu_6 = \frac{n}{4},\; \mu_8 = -\frac{17}{16}n,\; \mu_{10} = \frac{31}{4}n, \ldots$$

Example 3. What are the half-invariants of a complete binomial law of errors (the complete terms of $(p+q)^n$)? Here

$$e^{\frac{\mu_1}{1}\tau+\frac{\mu_2}{2!}\tau^2+\frac{\mu_3}{3!}\tau^3+\cdots} = \left(\frac{p+qe^{\tau}}{p+q}\right)^n$$

From this we obtain by differentiation with regard to τ

$$\mu_1+\frac{\mu_2}{1}\tau+\frac{\mu_3}{2!}\tau^2+\frac{\mu_4}{3!}\tau^3+\cdots = \frac{nqe^{\tau}}{p+qe^{\tau}},$$

by further differentiation

$$\mu_{i+1}+\frac{\mu_{i+2}}{1}\tau+\cdots = \frac{d^i\,\frac{nqe^{\tau}}{p+qe^{\tau}}}{d\tau^i},$$

putting $\tau=0$ we get

$$\mu_1 = \frac{nq}{p+q}$$

$$\mu_2 = \frac{npq}{(p+q)^2}$$

$$\mu_3 = \frac{npq(p-q)}{(p+q)^3}$$

$$\mu_4 = \frac{npq}{(p+q)^2}\left(\left(\frac{p-q}{p+q}\right)^2 - \frac{2pq}{(p+q)^2}\right)$$

$$\mu_5 = \frac{npq}{(p+q)^2}\left(\left(\frac{p-q}{p+q}\right)^3 - \frac{8pq(p-q)}{(p+q)^3}\right)$$

Example 4 A law of presumptive errors is given by its half-invariants forming a geometrical progression, $\lambda_r = ba^r$ Determine the several observations and their frequencies. Here the left hand side of the equation (18) is

$$s_0 e^{b\frac{a\tau}{\underline{|1}} + b\frac{(a\tau)^2}{\underline{|2}} + b\frac{(a\tau)^3}{\underline{|3}} + \cdots} = s_0 e^{-b} e^{be^{a\tau}},$$

but this is $= s_0 e^{-b}\left(1 + be^{a\tau} + \frac{b^2}{\underline{|2}} e^{2a\tau} + \frac{b^3}{\underline{|3}} e^{3a\tau} + \cdots\right)$ and has also the form of the right side of (18) Thus the observed values are $0, a, 2a, 3a, \ldots$, and the relative frequency of ra is $\frac{b^r}{\underline{|r}} = \varphi(r)$ This law of errors is nearly related to the binomial law, which can be considered as a product of two factors of this kind,

$$\frac{b^r}{\underline{|r}}\,\frac{d^{n-r}}{\underline{|n-r}} = \frac{1}{\underline{|n}}\,\beta_n(r)\,b^r d^{n-r}.$$

It is perhaps superior to the binomial law as a representative of some skew laws of errors

Example 5 A law of errors has the peculiarity that all half-invariants of odd order are $= 0$, while all even half-invariants are equal to each other, $\lambda_{2r} = 2a$ Show that all the observations must be integral numbers, and that for the relative frequencies

$$\varphi(0) = e^{-2a}\left(1 + \left(\frac{a}{\underline{|1}}\right)^2 + \left(\frac{a^2}{\underline{|2}}\right)^2 + \cdots\right)$$

$$\varphi(\pm r) = e^{-2a}\left(\frac{a^r}{\underline{|0}\,\underline{|r}} + \frac{a^{r+2}}{\underline{|1}\,\underline{|r+1}} + \frac{a^{r+4}}{\underline{|2}\,\underline{|r+2}} + \cdots\right).$$

Example 6 Determine the half-invariants of the law of presumptive errors for the irrational values in the table of a function, in whose computation fractions under $\frac{1}{2}$ have been rejected and those over $\frac{1}{2}$ replaced by 1.

$$\lambda_{2r+1} = 0,\ \lambda_2 = \tfrac{1}{12},\ \lambda_4 = -\tfrac{1}{120},\ \lambda_6 = \tfrac{1}{252},$$

§ 25. As a most general functional form of a continuous law of errors we have proposed (6)

$$\Phi(x) = k_0\varphi(x) - \frac{k_1}{\underline{|1}} D\varphi(x) + \frac{k_2}{\underline{|2}} D^2\varphi(x) - \frac{k_3}{\underline{|3}} D^3\varphi(x) + \cdots,$$

where $\varphi(x) = e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2}$

Now it is a very remarkable thing that we can express the half-invariants without any ambiguity as functions of the coefficients k_i, and vice versa

By (29) we get

$$s_0 e^{\frac{\lambda_1}{\underline{|1}}\tau + \frac{\lambda_2}{\underline{|2}}\tau^2 + \dots} = \int_{-\infty}^{+\infty} e^{o\tau}\left(k_0\varphi(o) - \frac{k_1}{\underline{|1}} D\varphi(o) + \frac{k_2}{\underline{|2}} D^2\varphi(o) \dots\right) do,$$

where $s_0 = nk_0\sqrt{2\pi}$. By means of the lemma

$$\int e^{o\tau} D^r\varphi(o)\, do = e^{o\tau}\left\{(D^{r-1}\varphi(o) - \tau D^{r-2}\varphi(o) + \dots + (-\tau)^{r-1}\varphi(o)\right\} + (-\tau)^r \int e^{o\tau}\varphi(o)\, do,$$

which is easily demonstrated for any $\varphi(o)$ by differentiating with regard to o only, we have in this particular case, where $\varphi(o)$ and every $D^r\varphi(o)$ is $= 0$, if $o = \pm\infty$,

$$\int_{-\infty}^{+\infty} e^{o\tau} D^r e^{-\frac{1}{2}\left(\frac{o-m}{n}\right)^2} do = (-\tau)^r \int_{-\infty}^{+\infty} e^{o\tau - \frac{1}{2}\left(\frac{o-m}{n}\right)^2} do = (-\tau)^r n\sqrt{2\pi}\, e^{m\tau + \frac{n^2}{2}\tau^2}.$$

Consequently, the relation between the half-invariants on one side and the coefficients k of the general functional law of errors on the other, is

$$k_0 e^{\frac{\lambda_1}{\underline{|1}}\tau + \frac{\lambda_2}{\underline{|2}}\tau^2 + \frac{\lambda_3}{\underline{|3}}\tau^3 + \dots} = \left(k_0 + \frac{k_1}{\underline{|1}}\tau + \frac{k_2}{\underline{|2}}\tau^2 + \frac{k_3}{\underline{|3}}\tau^3 \dots\right) e^{m\tau + \frac{1}{2}n^2\tau^2} \qquad (30)$$

If we write here $\lambda'_1 = \lambda_1 - m$ and $\lambda'_2 = \lambda_2 - n^2$, the computation of one set of constants by the other can, according to (17), be made by the formulæ (19) and (21). We substitute only in these the k_i for the s_i, and λ' or λ for μ.

It will be seen that the constants m and n, and the special typical law of errors to which they belong, are generally superfluous. This superfluity in our transformation may be useful in special cases for reasons of convergency, but in general it must be considered a source of vagueness, and the constants must be fixed arbitrarily.

It is easiest and most natural to put

$$m = \lambda_1 \text{ and } n^2 = \lambda_2$$

In this case we get $k_1 = 0$, $k_2 = 0$, $k_3 = k_0\lambda_3$, $k_4 = k_0\lambda_4$, $k_5 = k_0\lambda_5$, and further

$$k_6 = k_0(\lambda_6 + 10\lambda_3^2)$$
$$k_7 = k_0(\lambda_7 + 35\lambda_3\lambda_4)$$
$$k_8 = k_0(\lambda_8 + 56\lambda_3\lambda_5 + 35\lambda_4^2)$$

The law of the coefficients is explained by writing the right side of equation (30)

$$k_0 e^{m\tau + \frac{n^2}{2}\tau^2 + \log\left(k_0 + \frac{k_3}{\underline{|3}}\tau^3 + \frac{k_4}{\underline{|4}}\tau^4 + \dots\right) - \log k_0}$$

Expressed by half-invariants in this manner the explicit form of equation (6) is

$$\varphi(x) = \frac{s_0}{\sqrt{2\pi\lambda_2}} e^{-\frac{1}{2}\frac{(x-\lambda_1)^2}{\lambda_2}} \left\{ \begin{array}{l} 1 + \frac{\lambda_3}{6\lambda_2^3}\left((x-\lambda_1)^3 - 3\lambda_2(x-\lambda_1)\right) + \\ + \frac{\lambda_4}{24\lambda_2^4}\left((x-\lambda_1)^4 - 6\lambda_2(x-\lambda_1)^2 + 3\lambda_2^2\right) + \\ + \frac{\lambda_5}{120\lambda_2^5}\left((x-\lambda_1)^5 - 10\lambda_2(x-\lambda_1)^3 + 15\lambda_2^2(x-\lambda_1)\right) + \end{array} \right\} \quad (31)$$

VIII LAWS OF ERRORS OF FUNCTIONS OF OBSERVATIONS

§ 26 There is nothing inconsistent with our definitions in speaking of laws of errors relating to any group of quantities which, though not obtained by repeated observations, have the like property, namely, that repeated estimations of a single thing give rise, owing to errors of one kind or other, to multiple and slightly differing results which are *prima facie* equally valid The various forms of laws of actual errors are indeed only summary expressions for such multiplicity, and the transition to the law of presumptive errors requires, besides this, only that the multiplicity is caused by fixed but unknown circumstances, and that the values must be mutually independent in that sense that none of the circumstances have connected some repetitions to others in a manner which cannot be common to all Compare § 24, Example 6

It is, consequently, not difficult to define the law of errors for a function of *one* single observation. Provided only that the function is univocal, we can from each of the observed values $o_1, o_2 \ldots o_n$ determine the corresponding value of the function, and

$$f(o_1), f(o_2), \ldots f(o_n)$$

will then be the series of repetitions in the law of errors of the function, and can be treated quite like observations

With respect, however, to those forms of laws of errors which make use of the idea of frequency (probability) we must make one little reservation Even though o_i and o_k are different, we can have $f(o_i) = f(o_k)$, and in this case the frequencies must evidently be added together Here, however, we need only just mention this, and remark that the laws of errors when expressed by half-invariants or other symmetrical functions are not influenced by it

Otherwise the frequency is the same for $f(o_i)$ as for o_i, and therefore also the probability The ordinates of the curves of errors are not changed by observations with discontinuous values; but the abscissa o_i is replaced by $f(o_i)$, and likewise the argument in the functional law of errors, In continuous functions, on the other hand, it is the areas between corresponding ordinates which must remain unchanged

In the form of symmetrical functions the law of errors of functions of observations may be computed, and not only when we know all the several observed values, and can therefore compute, for each of them, the corresponding value of the function, and at last the symmetrical functions of the latter. In many and important cases it is sufficient if we know the symmetrical functions of the observations, as we can compute the symmetrical functions of the functions directly from these. For instance, if $f(o) = o^2$, for then the sums of the powers s'_n of the squares are also sums of the powers s_m of the observations, if only constantly $m = 2n$, $s'_0 = s_0$, $s'_1 = s_2$, $s'_2 = s_4$, etc.

§ 27. The principal thing is here a proposition as to laws of errors of the *linear functions* by half-invariants.

It is almost self-evident that if $o' = ao + b$

$$\left.\begin{aligned} \mu'_1 &= a\mu_1 + b \\ \mu'_2 &= a^2\mu_2 \\ \mu'_3 &= a^3\mu_3 \\ &\text{etc.} \\ \mu'_r &= a^r\mu_r \; (r > 1) \end{aligned}\right\} \qquad (32)$$

For the linear functions can always be considered as produced by the change of both zero and unity of the observations (Compare (24)).

However special the linear function $ao + b$ may be, we always in practice manage to get on with the formula (32). That we can succeed in this is owing to a happy circumstance, the very same as, in numerical solutions of the problems of exact mathematics, brings it about that we are but rarely, in the neighbourhood of equal roots, compelled to employ the formulæ for the solution of other equations than those of the first degree. Here we are favoured by the fact that we may suppose the errors in *good* observations to be small, so small — to speak more exactly — that we may generally in repetitions for each series of observations $o_1, o_2, \ldots o_n$ assign a number c, so near them all that the squares and products and higher powers of the differences

$$o_1 - c, \; o_2 - c, \; \ldots \; o_n - c$$

without any perceptible error may be left out of consideration in computing the function; i. e., these differences are treated like differentials. The differential calculus gives a definite method, in such circumstances, for transforming any function $f(o)$ into a linear one

$$f(o) = f(c) + f'(c) \; (o - c)$$

The law of errors then becomes

$$\left.\begin{aligned} \mu_1(f(o)) &= f(c) + f'(c)(\mu_1(o) - c) = f(\mu_1(o)) \\ \mu_r(f(o)) &= (f'(c))^r \mu_r(o) \end{aligned}\right\} \qquad (33)$$

But also by quite elementary means and easy artifices we may often transform functions into others of linear form If for instance $f(o) = \frac{1}{o}$, then we write

$$\frac{1}{o} = \frac{1}{c + (o - c)} = \frac{c - (o - c)}{c^2 - (o - c)^2} = \frac{1}{c} - \frac{1}{c^2}(o - c),$$

and the law of errors is then

$$\mu_1\left(\frac{1}{o}\right) = \frac{1}{c} - \frac{1}{c^2}(\mu_1(o) - c)$$

$$\mu_2\left(\frac{1}{o}\right) = \frac{1}{c^4}\mu_2(o)$$

$$\mu_r\left(\frac{1}{o}\right) = \frac{(-1)^r}{c^{2r}}\mu_r(o)$$

§ 28 With respect to *functions of two or more observed quantities* we may also, in case of repetitions, speak of laws of errors, only we must define more closely what we are to understand by repetitions For then another consideration comes in, which was out of the question in the simpler case It is still necessary for the idea of the law of errors of $f(o, o')$ that we should have, for each of the observed quantities o and o', a series of statements which severally may be looked upon as repetitions

$$o_1, \; o_2, \qquad o_n$$
$$o'_1, \; o'_2, \qquad o'_n$$

But here this is not sufficient Now it makes a difference if, among the special circumstances by o and o', there are or are not such as are common to observations of the different series We want a technical expression for this Here it is not appropriate only to speak of observations which are, respectively, dependent on one another or independent, we are led to mistake the partial dependence of observations for the functional dependence of exact quantities I shall propose to designate these particular interdependences of repetitions of different observations by the word "bond", which presumably cannot cause any misunderstanding

Among the repetitions of a single observation, no other bonds must be found than such as equally bind all the repetitions together, and consequently belong to the peculiarities of the method But while, for instance, several pieces cast in the same mould may be fair repetitions of one another, and likewise one dimension measured once on each piece, two or more dimensions measured on the same piece must generally be supposed to be bound together And thus there may easily exist bonds which, by community in a circumstance, as here the particularities in the several castings, bind some or all the repetitions of a series each to its repetition of another observation, and if observations thus connected are to enter into the same calculation, we must generally take these bonds into account. This, as a rule, can only be done by proposing a theory or hypothesis as to the

mathematical dependence between the observed objects and their common circumstance, and whether the number which expresses this is known from observation or quite unknown, the right treatment falls under those methods of adjustment which will be mentioned later on.

It is then in a few special cases only that we can determine laws of errors for functions of two or more observed quantities, in ways analogous to what holds good of a single observation and its functions.

If the observations $o, o', o'', \ldots$ which are to enter into the calculation of $f(o, o', o'', \ldots)$, are repeated in such a way that, in general, $o_i, o'_i, o''_i, \ldots$ of the i'th repetition are connected by a common circumstance, the same for each i, but otherwise without any other bonds, we can for each i compute a value of the function $y_i = f(o_i, o'_i, o''_i, \ldots)$, and laws of errors can be determined for this, in just the same way as for o separately. To do so we need no knowledge at all of the special nature of the bonds.

§ 29. If, on the contrary, there is no bond at all between the repetitions of the observations $o, o', o'', \ldots$ — and this is the principal case to which we must try to reduce the others — then we must, in order to represent all the equally valid values of $y = f(o, o', o'', \ldots)$, herein combine every observed value for o with every one for o', for o'', etc., and all such values of y must be treated analogously to the simple repetitions of one single observed quantity. But while it may here easily become too great a task to compute y for each of the numerous combinations, we shall in this case be able to compute y's law of errors by means of the laws of errors for $o, o', o'' \ldots$

Concerning this a number of propositions might be laid down; but one of them is of special importance and will be almost sufficient for us in what follows, viz., that which teaches us to determine the law of errors for the sum O of the observed quantities o and o'.

If the law of errors is given in the form of relative frequencies or probabilities, $\varphi(o)$ for o and $\psi(o')$ for o', then it is obvious that the product $\varphi(o)\psi(o')$ must be the frequency of the special sum $o + o'$.

In the calculus of probabilities, we shall consider this form more closely, and there some cases of bound observations will find their solution; here we shall confine ourselves to the treatment of the said case with half-invariants.

If o occurs with the observed values

$$o_1, o_2, \ldots o_m$$

and o' with

$$o'_1, o'_2, \ldots o'_n,$$

then by the mn repetitions of the operation $O = o + o'$ we get

$$\begin{array}{lll} o_1+o'_1, & o_1+o'_2, \ldots & o_1+o'_n, \\ o_2+o'_1, & o_2+o'_2, \ldots & o_2+o'_n, \\ \ldots & & \ldots \\ o_m+o'_1, & o_m+o'_2, \ldots & o_m+o'_n \end{array}$$

Indicating by M_r the half-invariants of the sum $O = o+o'$, we get by (18)

$$m \cdot n\, e^{\frac{M_1}{\underline{1}}\tau+\frac{M_2}{\underline{2}}\tau^2+\frac{M_3}{\underline{3}}\tau^3+\ldots} = \Sigma e^{(o+o')\tau} = (e^{o_1\tau}+\ldots e^{o_m\tau})(e^{o'_1\tau}+\ldots e^{o'_n\tau})$$

where m and n are the numbers of repetitions of o and o'. Consequently, if μ_r represent the half-invariants of o, and μ'_r of o', we get

$$e^{\frac{M_1}{\underline{1}}\tau+\frac{M_2}{\underline{2}}\tau^2+\ldots} = e^{\frac{\mu_1}{\underline{1}}\tau+\frac{\mu_2}{\underline{2}}\tau^2+\ldots} \cdot e^{\frac{\mu'_1}{\underline{1}}\tau+\frac{\mu'_2}{\underline{2}}\tau^2+\ldots}$$

and finally

$$\left.\begin{array}{l} M_1 = \mu_1+\mu'_1 \\ \ldots \\ M_r = \mu_r+\mu'_r \end{array}\right\} \qquad (34)$$

Employing the equation (17) instead of (18) we can also obtain fairly simple expressions for the sums of powers of $(o+o')$ analogous to the binomial formula. But the extreme simplicity of (34) renders the half-invariants unrivalled as the most suitable symmetrical functions and the most powerful instrument of the theory of observations.

More generally, for every linear function of observations not connected by any bond,

$$O = a+bo+co'+\ldots do'',$$

we obtain in the same manner and by (32)

$$\left.\begin{array}{l} M_1(o) = a+b\mu_1+c\mu'_1+\ldots+d\mu'''_1 \\ M_2(o) = \quad b^2\mu_2+c^2\mu'_2+\ldots+d^2\mu'''_2 \\ \ldots \\ M_r(o) = \quad b^r\mu_r+c^r\mu'_r+\ldots+d^r\mu'''_r \\ \qquad\qquad r>1 \end{array}\right\} \qquad (35)$$

When the errors of observation are sufficiently small, we shall also here generally be able to give the most different functions a linear form. In consequence of this, the propositions (34) and (35) acquire an almost universal importance, and afford nearly the whole necessary foundation for the theory of the laws of errors of functions.

Example 1. Determine the square of the mean error for differences of the n'th order of equidistant tabular values, between which there is no bond, the square of the mean error for every value being $= \lambda_2$.

$$\lambda_2(\Delta^1) = \lambda_2(o_1 - o_0) = 2\lambda_2$$
$$\lambda_2(\Delta^2) = \lambda_2(o_2 - 2o_1 + o_0) = 6\lambda_2$$
$$\lambda_2(\Delta^3) = \lambda_2(o_3 - 3o_2 + 3o_1 - o_0) = 20\lambda_2$$
$$\lambda_2(\Delta^4) = \lambda_2(o_4 - 4o_3 + 6o_2 - 4o_1 + o_0) = 70\lambda_2$$

. . .

$$\lambda_2(\Delta^n) = \frac{2}{1}\cdot\frac{6}{2}\cdot\frac{10}{3}\cdot\frac{14}{4} \cdots \frac{4n-2}{n}\lambda_2$$

Example 2. By the observation of a meridional transit we observe two quantities, viz the time, t, when a star is covered behind a thread, and the distance, f, from the meridian at that instant But as it may be assumed that the time and the distance are not connected by a bond, and as the speed of the star is constant and proportional to the known value $\sin p$ (p = polar distance), we always state the observation by the one quantity, the time when the very meridian is passed, which we compute by the formula $o = t + f \operatorname{cosec} p$.

The mean error is
$$\lambda_2(o) = \lambda_2(t) + \operatorname{cosec}^2 p\,\lambda_2(f)$$

Example 3 A scale is constructed by making marks on it at regular intervals, in such a way that the square of the mean error on each interval is $= \lambda_2$.

To measure the distance between two objects, we determine the distance of each object from the nearest mark, the square of the mean error of this observation being $= \lambda'_2$. How great is the mean error in a measurement, by which there are n intervals between the marks we use?
$$\lambda_2(\text{length}) = n\lambda_2 + 2\lambda'_2$$

Example 4 Two points are supposed to be determined by bond-free and equally good ($\lambda_2 = 1$) measurements of their rectangular co ordinates The errors being small in proportion to the distance, how great is the mean error in the distance Δ?
$$\lambda_2(\Delta) = 2$$

Example 5. Under the same suppositions, what is the mean error in the inclination to the x-axis?
$$\lambda_2(E) = \frac{2}{\Delta^2}$$

Example 6 Having three points in a plane determined in the same manner by their rectangular co-ordinates (x_1, y_1), (x_2, y_2), (x_3, y_3), find the mean error of the angle at the point (x_1, y_1)
$$\lambda_2(V) = \frac{\Delta_1^2 + \Delta_2^2 + \Delta_3^2}{\Delta_2^2\,\Delta_3^2},$$

$\Delta_1, \Delta_2, \Delta_3$ being the sides of the triangle; Δ_1 opposite to (x_1, y_1).

Examples 7 and 8. Find the mean errors in determinations of the areas of a triangle and a plane quadrangle.

$$\lambda_2(\text{triangle}) = \tfrac{1}{4}(\Delta_1^2 + \Delta_2^2 + \Delta_3^2); \quad \lambda_2(\text{quadrangle}) = \tfrac{1}{4}\left(\Delta_{13}^2 + \Delta_{24}^2\right)$$

§ 90. Non-linear functions of more than one argument present very great difficulties. Even for integral rational functions no general expression for the law of errors can be found. Nevertheless, even in this case it is possible to indicate a method for computing the half-invariants of the function by means of those of the arguments. To do so it seems indispensable to transform the laws of errors into the form of systems of sums of powers. If $O = f(o, o', \dots o^{(m)})$ be integral and rational, both it and its powers O^r can be written as sums of terms of the standard form $\Sigma k o^a o'^b \dots o^{(m)d}$, and for every such term the sum resulting from the combination of all repetitions is $k s_a s'_b \dots s_d^{(m)}$ (including the cases where a or b or d may be $= 0$), $s_c^{(i)}$ being the sum of all c^{th} powers of the repetitions of $o^{(i)}$. Thus if S_r indicates the sum of the r^{th} powers of the function O, we get

$$S_r = \Sigma k s_a s'_b \dots s_d^{(m)}.$$

Of course, this operation is only practicable in the very simplest cases.

Example 1. Determine the mean value and mean deviation of the product $oo' = O$ of two observations without bonds. Here $S_0 = s_0 s'_0$ and generally $S_r = s_r s'_r$, consequently the mean value $M_1 = \mu_1 \mu'_1$ and

$$M_2 = \mu_2 \mu'_2 + \mu_2 \mu_1'^2 + \mu'_2 \mu_1^2.$$

M_3 already takes the cumbersome form

$$M_3 = \mu_3 \mu'_3 + \mu_3 \mu'_1 (3\mu'_2 + \mu_1'^2) + \mu'_3 \mu_1 (3\mu_2 + \mu_1^2) + 6\mu_2 \mu_1 \mu'_2 \mu'_1.$$

Example 2. Express exactly by the half-invariants of the co-ordinates the mean value and the mean deviation of the square of the distance $r^2 = x^2 + y^2$, if x and y are observed without bonds. Here

$$s_0(r^2) = s_0(x) s_0(y)$$
$$s_1(r^2) = s_2(x) s_0(y) + s_0(x) s_2(y)$$
$$s_2(r^2) = s_4(x) s_0(y) + 2 s_2(x) s_2(y) + s_0(x) s_4(y)$$

and

$$\mu_1(r^2) = \mu_2(x) + (\mu_1(x))^2 + \mu_2(y) + (\mu_1(y))^2$$
$$\mu_2(r^2) = \mu_4(x) + 4\mu_3(x)\mu_1(x) + 2(\mu_2(x))^2 + 4\mu_2(x)(\mu_1(x))^2 +$$
$$+ \mu_4(y) + 4\mu_3(y)\mu_1(y) + 2(\mu_2(y))^2 + 4\mu_2(y)(\mu_1(y))^2.$$

§ 91. The most important application of proposition (85) is certainly the determination of the law of errors of the mean value itself. The mean value

$$\mu_1 = \frac{1}{n}(o_1 + o_2 + \dots o_n)$$

is, we know, a linear function of the observed values, and we may treat the law of errors for μ_1 according to the said proposition, not only where we look upon $o_1, \ldots o_m$ as perfectly unconnected, but also where we assume that they result from repetitions made according to the same method For, just like such repetitions, $o_1, \ldots o_m$ must not have any other circumstances in common as connecting bonds than such as bind them all and characterise the method

As the law of presumptive errors of o_1 is just the same as for $o_2, \ldots o_m$, with the known half-invariants $\lambda_1, \lambda_2, \ldots \lambda_r, \ldots$, we get according to (36)

$$\left.\begin{aligned} \lambda_1(\mu_1) &= \frac{1}{m}(\lambda_1 + \ldots + \lambda_1) = \lambda_1, \\ \lambda_2(\mu_1) &= \frac{1}{m^2}(\lambda_2 + \ldots + \lambda_2) = \frac{1}{m}\lambda_2 \\ \text{and in general} \quad \lambda_r(\mu_1) &= m^{1-r}\lambda_r. \end{aligned}\right\} \quad (37)$$

While, consequently, the presumptive mean of a mean value for m repetitions is the presumptive mean itself, the mean error on the mean value μ_1 is reduced to $\frac{1}{\sqrt{m}}$ of the mean error on the single observation. When the number m is large, the formation of mean values consequently reduces the uncertainty considerably; the reduction, however, is proportionally greater with small than with large numbers While already 4 repetitions bring down the uncertainty to half of the original, 100 repetitions are necessary in order to add one significant figure, and a million to add 3 figures to those due to the single observation.

The higher half-invariants of μ_1 are reduced still more If the λ_3, λ_4, etc., of the single observation are so large that the law of errors cannot be called typical, no very great numbers of m will be necessary to realise the conditions $\lambda_3(\mu_1) = 0 = \lambda_4(\mu_1)$ with an approximation that is sufficient in practice It ought to be observed that this reduction is not only absolute, but it holds good also in relation to the corresponding power of the mean error $\sqrt{\lambda_2(\mu_1)^2}$; for (37) gives

$$\lambda_r(\mu_1) \cdot (\lambda_2(\mu_1))^{-\frac{r}{2}} = m^{1-\frac{r}{2}} \cdot (\lambda_r \lambda_2^{-\frac{r}{2}}),$$

which, for instance when $m = 4$, shows that the deviation of λ_3 from the typical form which appears by means of only 4 repetitions, is halved, that of λ_4 is divided by 4, that of λ_5 is divided by 8, etc. This shows clearly the reason why we *attach great importance to the typical form for the law of errors* and make arrangements to abide by it in practice. For it appears now that *we possess in the formation of mean values a means of making the laws of errors typical*, even where they were not so originally Therefore the standard rule for all practical observations is this Take care not to neglect any opportunities of

repeating observations and parts of observations, so that you can directly form the mean values which should be substituted for the observed results, and this is to be done especially in the case of observations of a novel character, or with peculiarities which lead us to doubt whether the law of errors will be typical

This remarkable property is peculiar, however, not to the mean only, but also, though with less certainty, to any linear function of several observations, provided only the coefficient of any single term is not so great relatively to the corresponding deviation from the typical form that it throws all the other terms into the shade From (95) it is seen that, if the laws of errors of all the observations $o, o', \ldots o^{(n)}$ are typical, the law of errors for any of their linear functions will be typical too And if the laws of errors are not typical, then that of the linear function will deviate relatively less than any of the observations $o, o', \ldots o_n$.

To avoid unnecessary complication we represent two terms of the linear function simply by o and o' The deviation from the typical zero, which appears in the r^{th} half-invariants ($r > 2$), measured by the corresponding power of the mean error, will be less for $O = o + o'$ than for the most discrepant of the terms o and o'

The inequation

$$\frac{\lambda_r^2}{\lambda_2^r} \geqq \frac{\lambda_r'^2}{\lambda_2'^r}$$

says only that, if the laws of errors for o and o' deviate unequally from the typical form, it is the law of errors for o that deviates most But this involves

$$\left(\frac{\lambda_2'}{\lambda_2}\right)^r \geqq \left(\frac{\lambda_r'}{\lambda_r}\right)^2$$

or more briefly

$$T^r \geqq R^2,$$

where T is positive, $r > 2$

When we introduce a positive quantity U, so that

$$T^r = U^2 \geqq R^2,$$

it is evident that $(U+1)^2 \geqq (R+1)^2$, and it is easily demonstrated that $(T+1)^r > (U+1)^2$.

Remembering that $x + x^{-1} \geqq 2$, if $x > 0$, we get by the binomial formula

$$\left(U^{\frac{1}{r}} + U^{-\frac{1}{r}}\right)^r \geqq U + U^{-1} + 2^r - 2 > (U^{\frac{1}{2}} + U^{-\frac{1}{2}})^2$$

Consequently

$$(T+1)^r > (U+1)^2 \geqq (R+1)^2$$

or

$$\left(\frac{\lambda_2 + \lambda_2'}{\lambda_2}\right)^r > \left(\frac{\lambda_r + \lambda_r'}{\lambda_r}\right)^2$$

and

$$\frac{\lambda_r^2}{\lambda_2^r} > \frac{(\lambda_r + \lambda_r')^2}{(\lambda_2 + \lambda_2')^r} = \frac{(\lambda_r(O))^2}{(\lambda_2(O))^r},$$

but this is the proposition we have asserted, for the extension to any number of terms causes no difficulty

But if it thus becomes a general law that the law of errors of linear functions must more or less approach the typical form, the same must hold good also of all moderately complex observations, such as those whose errors arise from a considerable number of sources The expression "source of errors" is employed to indicate circumstances which undeniably influence the result, but which we have been obliged to pass over as unessential If we imagined those circumstances transferred to the class of essential circumstances, and substantiated by subordinate observations, that which is now counted an observation would occur as a function, into which the subordinate observations enter as independent variables, and as we may assume, in the case of good observations, that the influence of each single source of errors is small, this function may be regarded as linear The approximation to typical form which its law of errors would thus show, if we knew the laws of errors of the sources of error, cannot be lost, simply because we, by passing them over as unessential, must consider the sources of error in the compound observation as unknown Moreover, we may take it for granted that, in systematically arranged observations, every such source of error as might dominate the rest will be the object of special investigation and, if necessary, will be included among the essential circumstances or removed by corrective calculations The result then is that great deviations from the typical form of the law of errors are rare in practice

§ 32 It is of interest, of course, also to acquire knowledge of the laws of errors for the determinations of μ_2 and the higher half-invariants as functions of a given number of repeated observations

Here the method indicated in § 30 must be applied But though the symmetry of those functions and the identity of the laws of presumptive errors for $o_1, o_2, \ldots o_m$ afford very essential simplifications, still that method is too difficult Not even for μ_2 have I discovered the general law of errors In my *"Almindelig Iagttagelseslære"*, Kobenhavn 1889, I have published tables up to the eighth degree of products of the sums of powers $s_p s_q \ldots$, expressed by sums of terms of the form o^i, o^{ij}, o^{ijk}, these are here directly applicable In W Fiedler *"Elemente der neueren Geometrie und der Algebra der binären Formen"*, Leipzig 1862, tables up to the 10th degree will be found Their use is more difficult, because they require the preliminary transformation of the s_p to the coefficients a_p of the rational equations § 21 There are such tables also in the *Algebra* by Meyer Hirsch, and Cayley has given others in the Philosophical Transactions 1857 (Vol 147,

p 48) I have computed the four principal half-invariants of μ_2

$$
\left.
\begin{aligned}
m\lambda_1(\mu_2) &= (m-1)\lambda_2 \\
m^3\lambda_2(\mu_2) &= (m-1)^2\lambda_4 + 2m(m-1)\lambda_2^2 \\
m^5\lambda_3(\mu_2) &= (m-1)^3\lambda_6 + 12m(m-1)^2\lambda_4\lambda_2 + 4m(m-1)(m-2)\lambda_3^2 + \\
&\quad + 8m^2(m-1)\lambda_2^3 \\
m^7\lambda_4(\mu_2) &= (m-1)^4\lambda_8 + 24m(m-1)^3\lambda_6\lambda_2 + 32m(m-1)^2(m-2)\lambda_5\lambda_3 + \\
&\quad + 8m(m-1)(4m^2-9m+6)\lambda_4^2 + 144m^2(m-1)^2\lambda_4\lambda_2^2 + \\
&\quad + 96m^2(m-1)(m-2)\lambda_3^2\lambda_2 + 48m^3(m-1)\lambda_2^4
\end{aligned}
\right\} \quad (38)
$$

Here m is the number of repetitions

Of μ_3 and μ_4 only the mean values and the mean errors have been found

$$
\left.
\begin{aligned}
m^2\lambda_1(\mu_3) &= (m-1)(m-2)\lambda_3, \\
m^5\lambda_2(\mu_3) &= (m-1)^2(m-2)^2\lambda_6 + 9m(m-1)(m-2)^2(\lambda_4\lambda_2 + \lambda_3^2) + \\
&\quad + 6m^2(m-1)(m-2)\lambda_2^3,
\end{aligned}
\right\} \quad (39)
$$

and

$$
\left.
\begin{aligned}
m^3\lambda_1(\mu_4) &= (m-1)(m^2-6m+6)\lambda_4 - 6m(m-1)\lambda_2^2 \\
m^7\lambda_2(\mu_4) &= (m-1)^2(m^2-6m+6)^2\lambda_8 + \\
&\quad + 8m(m-1)(m^2-6m+6)(2m^2-15m+15)\lambda_6\lambda_2 + \\
&\quad + 18m(m-1)(m-2)(m-4)(m^2-6m+6)\lambda_5\lambda_3 + \\
&\quad + 2m(m-1)(17m^4-204m^3+852m^2-1404m+828)\lambda_4^2 + \\
&\quad + 24m^2(m-1)(3m^3-38m^2+150m-168)\lambda_4\lambda_2^2 + \\
&\quad + 144m^2(m-1)(m-2)(m-4)(m-5)\lambda_3^2\lambda_2 + \\
&\quad + 24m^3(m-1)(m^2-6m+24)\lambda_2^4
\end{aligned}
\right\} \quad (40)
$$

Further I know only that

$$
m^4\lambda_1(\mu_5) = (m-1)(m-2)\{(m^2-12m+12)\lambda_5 - 60m\lambda_3\lambda_2\}, \quad (41)
$$

$$
\begin{aligned}
m^5\lambda_1(\mu_6) &= (m-1)(m^4-30m^3+150m^2-240m+120)\lambda_6 - \\
&\quad - 30m(m-1)(7m^2-30m+30)\lambda_4\lambda_2 - \\
&\quad - 10m(m-1)(m-2)(3m-8)\lambda_3^2 - \\
&\quad - 30m^2(m-1)(m-6)\lambda_2^3,
\end{aligned} \quad (42)
$$

$$
\begin{aligned}
m^6\lambda_1(\mu_7) &= (m-1)(m-2)(m^4-60m^3+420m^2-720m+360)\lambda_7 - \\
&\quad - 630m(m-1)(m-2)(m^2-8m+8)\lambda_5\lambda_2 - \\
&\quad - 210m(m-1)(m-2)(7m^2-48m+60)\lambda_4\lambda_3 - \\
&\quad - 1260m^2(m-1)(m-2)(m-10)\lambda_3\lambda_2^2,
\end{aligned} \quad (43)
$$

$$m^7\lambda_1(\mu_8) = (m-1)(m^6-120m^5+1806m^4-8400m^3+10800m^2-15120m+5040)\lambda_8 -$$
$$- 56m(m-1)(31m^4-540m^3+2340m^2-3600m+1800)\lambda_6\lambda_2 -$$
$$- 1680m(m-1)(m-2)(3m^3-40m^2+120m-96)\lambda_5\lambda_3 -$$
$$- 70m(m-1)(49m^4-720m^3+3108m^2-5400m+3240)\lambda_4^2 -$$
$$- 840m^2(m-1)(7m^3-150m^2+576m-540)\lambda_4\lambda_2^2 -$$
$$- 10080m^2(m-1)(m-2)(m^2-16m+40)\lambda_3^2\lambda_2 -$$
$$- 840m^3(m-1)(m^2-30m+90)\lambda_2^4 \qquad (44)$$

Some λ_1's of products of the μ_2, μ_3, and μ_4 present in general the same characteristics as the above formulæ. The most prominent of these characteristics are

1) It is easily explained that λ_1 is only to be found in the equation $\lambda_1(\mu_1) = \lambda_1$; indeed no other half-invariant than the mean value can depend on the zero of the observations. In my computations this characteristic property has afforded a system of multiple checks of the correctness of the above results.

2) All mean $\lambda_1(\mu_r)$ are functions of the 0^{th} degree with regard to m, all squares of mean errors $\lambda_2(\mu_r)$, are of the $(-1)^{st}$ degree, and generally each $\lambda_s(\mu_r)$ is a function of the $(1-s)^{th}$ degree, in perfect accordance with the law of large numbers.

3) The factor $m-1$ appears universally as a necessary factor of $\lambda_s(\mu_r)$, if only $r > 1$. If r is an odd number, even the factor $m-2$ appears, and, likewise, if r is an even number, this factor is constantly found in every term that is multiplied by one or more λ's with odd indices. No obliquity of the law of errors can occur unless at least three repetitions are under consideration.

4) Many particulars indicate these functions as compounds of factorials $(m-1)(m-2)\ldots(m-r)$ and powers of m.

If, supposing the presumed law of errors to be typical, we put $\lambda_3 = \lambda_4 = \ldots = 0$, then some further inductions can be made. In this case the law of errors of μ_2 may be

$$e^{\frac{\lambda_1(\mu_2)}{|1}\tau + \frac{\lambda_2(\mu_2)}{|2}\tau^2 + \cdots} = \left(1-\frac{2\lambda_2\tau}{m}\right)^{\frac{1-m}{2}} = \int_{-\infty}^{+\infty}\varphi(o)e^{o\tau}do \qquad (45)$$

As to the squares of mean errors of μ_r we get under the same supposition

$$\left.\begin{aligned}
\lambda_2(\mu_1) &= \tfrac{1}{m}\lambda_2 \\
\lambda_2(\mu_2) &= \tfrac{2}{m}\lambda_2^2 \\
\lambda_2(\mu_3) &= \tfrac{6}{m}\lambda_2^3 \\
\lambda_2(\mu_4) &= \tfrac{24}{m}\lambda_2^4, \\
\lambda_2(\mu_r) &= \tfrac{|r}{m}\lambda_2^r
\end{aligned}\right\} \qquad (46)$$

indicating that generally

This proposition is of very great interest If we have a number m of repetitions at our disposal for the computation of a law of actual errors, then it will be seen that the relative mean errors of $\mu_1, \mu_2, \mu_3 \ldots \mu_r$ are by no means uniform, but increase with the index r If m is large enough to give us μ_1 precisely and μ_2 fairly well then μ_3 and μ_4 can be only approximately indicated, and the higher half-invariants are only to be guessed, if the repetitions are not counted by thousands or millions.

As all numerical coefficients in $\lambda_2(\mu_r)$ increase with r, almost in the same degree as the coefficients 1, 2, 6, and 24 of λ_2^r, we must presume that the law of increasing uncertainty of the half-invariants has a general character.

We have hitherto been justified in speaking of the principal half-invariants as the complete collection of the μ_r's or λ_r's with the lowest indices, considering a complete series of the first m half-invariants to be necessary to an unambiguous determination of a law of errors for m repetitions

We now accept that principle as a system of relative rank of the half-invariants with increasing uncertainty and consequently with a decreasing importance of the half-invariants with higher indices.

We need scarcely say that there are some special exceptions to this rule For instance if $\lambda_4 = -\lambda_2^2$, as in alternative experiments with equal chances for and against (pitch and toss), then $\lambda_2(\mu_2)$ is reduced to $= \frac{2(m-1)}{m^3}\lambda_2^2$, which is only of the $(-2)^{nd}$ order

§ 33 Now we can undertake to solve the main problem of the theory of observations, the transition from laws of actual errors to those of presumptive errors. Indeed this problem is not a mathematical one, but it is eminently practical To reason from the actual state of a finite number of observations to the law governing infinitely numerous presumed repetitions is an evident trespass, and it is a mere attempt at prophecy to predict, by means of a law of presumptive errors, the results of future observations

The struggle for life, however, compels us to consult the oracles. But the modern oracles must be scientific, particularly when they are asked about numbers and quantities, mathematical science does not renounce its right of criticism We claim that confusion of ideas and every ambiguous use of words must be carefully avoided; and the necessary act of will must be restrained to the acceptation of fixed principles, which must agree with the law of large numbers

It is hardly possible to propose more satisfactory principles than the following.

The mean value of all available repetitions can be taken directly, without any change, as an approximation to the presumptive mean.

If only one observation without repetition is known, it must itself, consequently, be considered an approximation to the presumptive mean value

The solitary value of any symmetrical and univocal function of repeated observations

must in the same way, as an *isolated observation*, be considered *the* presumptive mean of this function, for instance $\mu_r = \lambda_1(\mu_r)$

Thus, from the equations 37—41, we get by m repetitions

$$\left.\begin{aligned}
\lambda_1 &= \mu_1 \\
\lambda_2 &= \frac{m}{m-1}\mu_2 \\
\lambda_3 &= \frac{m^2}{(m-1)(m-2)}\mu_3 \\
\lambda_4 &= \frac{m^3}{(m-1)(m^2-6m+6)}\left(\mu_4 + \frac{6}{m-1}\mu_2^2\right) \\
\lambda_5 &= \frac{m^4}{(m-1)(m-2)(m^2-12m+12)}\left(\mu_5 + \frac{60}{m-1}\mu_3\mu_2\right),
\end{aligned}\right\} \qquad (47)$$

as to λ_6, λ_7, λ_8 it is preferable to use the equations 42—44 themselves, putting only $\lambda_1(\mu_6) = \mu_6$, $\lambda_1(\mu_7) = \mu_7$, and $\lambda_1(\mu_8) = \mu_8$

Inversely, if the presumptive law of errors is known in this way, or by adoption of any theory or hypothesis, we *predict the future observations, or functions of observations, principally by computing their presumptive mean values* These predictions however, though univocal, are never to be considered as exact values, but only as the first and most important terms of laws of errors

If necessary, we complete our predictions with the mean errors and higher half-invariants, computed for the predicted functions of observations by the presumed law of errors, which itself belongs to the single observations These supplements may often be useful, nay necessary, for the correct interpretation of the prediction The ancient oracles did not release the questioner from thinking and from responsibility, nor do the modern ones; yet there is a difference in the manner. If the crossing of a desert is calculated to last 20 days, with a mean error of one day, then you would be very unwise, to be sure, if you provided for exactly 20 days, by so doing you incur as great a probability of dying as of living. Even with provisions for 21 days the journey is evidently dangerous But if you can carry with you provisions for 23—25 days, the undertaking may be reasonable Your life must be at stake to make you set out with provisions for only 17 days or less.

In addition to the uncertainty provided against by the presumptive law of error, the prediction may be vitiated by the uncertainty of the data of the presumptive law itself When this law has resulted from purely theoretical speculation, it is always impossible to calculate its uncertainty It may be quite exact, or partially or, absolutely false, we are left to choose between its admission and its rejection, as long as no trial of the prediction by repeated observations has given us a corresponding law of actual errors, by which it can be improved on

If the law of presumptive errors has been computed by means of a law of actual errors, we can, according to (37), employ the values λ_1, λ_2, and the number m of actual observations for the determination of $\lambda_r(\mu_1)$. In this case the complete half-invariants of a predicted single observation are given analogously to the law of errors of the sum *of two bondless observations by*

$$\lambda_1$$

$$\lambda_2 + \lambda_2(\mu_1)$$

$$\lambda_r + \lambda_r(\mu_1)$$

Though we can in the same way compute the uncertainties of λ_1, λ_2, and λ_3, it is far more difficult, or rather impossible, to make use of these results for the improvement *of general predictions*

Of the higher half-invariants we can very seldom, if ever, get so much as a rough estimate by the method of laws of actual errors The same reasons that cause this difficulty, render it a matter of less importance to obtain any precise determination Therefore *the general rule of the formation of good laws of presumptive errors* must be

1. In determining λ_1 and λ_2, rely almost entirely upon the actual observed values

2 As to the half-invariants with high indices, say from λ_6 upwards, rely as exclusively upon theoretical considerations

3 Employ the indications obtainable by actual observed values for the intermediate half-invariants as far as possible when you have the choice between the theories in (2)

From what is said above of the properties of *the typical law of errors*, it is evident that no other theory can fairly rival it in the multiplicity and importance of applications. It is not only constantly applied when λ_3, λ_4, and λ_5 are proved to be very small, but it is used almost universally as long as the deviations are not very conspicuous In these cases also great efforts will be made to reduce the observations to the typical form by *modifying the methods* or by substituting means of many observed values *instead of the* non-typical single observations The preference for the typical observations is intensified by the difficulty of establishing an entirely correct method of adjustment (see the following chapters) of observations which are not typical

In those particular cases where λ_3 or λ_4 or λ_5 cannot be *regarded* as small, the theoretical considerations (proposition 2 above) as to λ_6 and the higher half-invariants ought not to result in putting the latter $= 0$ As shown in "*Videnskabernes Selskabs Oversigter*", 1899, p 140, such laws of errors correspond to divergent series or imply the existence of *imaginary observations* *The coefficients* k_r *of the functional* law of *errors* (equation (0))

have this merit in preference to the half-invariants, that no term implies the existence of any other

This series

$$\theta(x) = k_0\varphi(x) - \frac{k_3}{|\underline{3}} D^3\varphi(x) + \frac{k_4}{|\underline{4}} D^4\varphi(x) - \frac{k_5}{|\underline{5}} D^5\varphi(x) + \frac{k_6}{|\underline{6}} D^6\varphi(x) \quad ,$$

where $\varphi(x) = \varepsilon^{-\frac{(x-\lambda_1)^2}{2\lambda_2}}$ (the direct expression (31) is found p. 35), is therefore recommended as perhaps the best general expression for non-typical laws of errors The functional form of the law of errors has here, and in every prediction of future results, the advantage of showing directly the probabilities of the different possible values.

The skew and other non-typical laws of errors seem to have some very interesting applications to biological observations, especially to the variations of species The scientific treatment of such variations seems altogether to require a methodical use of the notion of laws of errors Mr K Pearson has given a series of skilful computations of biological and other similar laws of errors (*Contributions to the Math Theory of Evolution, Phil. Trans* V 186, p 343). Here he makes very interesting efforts to develop the refractory binomial functions into a basis for the treatment of skew laws of errors But there are evidently no natural links between these functions and the biological problems, and the above formulæ (31) will prove to be easier and more powerful instruments In cases of very abnormal or discontinuous laws of errors, more refined methods of adjustment are required

Example 1 From the 500 experiments given in § 14 are to be calculated the presumptive half-invariants up to λ_5 and by (31) the frequencies of the special events out of a number of $s_1 = 500$ new repetitions. You will find $\lambda_1 = 11{\cdot}86$, $\lambda_2 = 4{\cdot}1647$, $\lambda_3 = 4{\cdot}708$, $\lambda_4 = 3{\cdot}895$, and $\lambda_5 = -26{\cdot}046$. A comparison of the computed frequencies with the observed ones gives:

	Frequency		
Events	computed	observed	o−c
4	0·0	0	− 0·0
5	−0·1	0	+ 0·1
6	−0·3	0	+ 0·3
7	1·6	3	+ 1·4
8	12·3	7	− 5·3
9	39·6	35	− 4·6
10	78·2	101	+22·8
11	104·1	89	−15·1
12	97·7	94	− 3·7
13	69·4	70	+ 0·6
14	42·8	46	+ 3·2

Events	Frequency computed	Frequency observed	o−c
15	26.7	30	+ 3.3
16	16.0	15	− 1.0
17	8.0	4	− 4.0
18	3.0	5	+ 2.0
19	0.8	1	+ 0.2
20	0.2	0	− 0.2
21	0.0	0	0.0

Example 2. Determine the law of errors by experiments with alternative results, either "yes" observed m times and every time indicated by 1, or "no" observed n times and indicated by 0. What is the square of the mean error for the single experiment?

$$\lambda_2 = \frac{mn}{(m+n)(m+n-1)},$$

for the probability determined by the whole series?

$$\lambda_2(\mu_1) = \frac{mn}{(m+n)^2 (m+n-1)},$$

and for the frequency of "yes" in the $m+n$ experiments?

$$\lambda_2(s_1) = \frac{mn}{m+n-1}.$$

§ 34. If observations are made and repeated, although their presumptive mean value is previously known, exactly or very accurately, the law of presumptive errors of the half-invariants $\mu_2, \mu_3 \ldots$ must be computed by reducing the zero of the observation to the known λ_1. Putting thus $s_1 = 0$ and $\mu_1 = 0$ in the equations (19) and (21) we obtain in analogy to (38)—(41) the following modified equations, the number of repetitions being $= m$,

$$\left.\begin{aligned}
\lambda_1(\mu_2) &= \mu_2 = \lambda_2 \\
\lambda_1(\mu_3) &= \mu_3 = \lambda_3 \\
\lambda_1(\mu_4) &= \mu_4 = \frac{m-3}{m}\lambda_4 - \frac{6}{m}\lambda_2^2 \\
\lambda_1(\mu_5) &= \mu_5 = \frac{m-10}{m}\lambda_5 - \frac{90}{m}\lambda_3\lambda_2
\end{aligned}\right\} \qquad (48)$$

From the first of these equations we deduce the very important principle, that every mean of the squares of differences between repeated bond-free observations and their presumptive mean value is approximately equal to the square of the mean error

$$\frac{\Sigma(o-\lambda_1)^2}{m} = \lambda_2 \qquad (49)$$

Consequently, for any isolated observed value we must expect that

$$(o - \lambda_1(o))^2 = \lambda_2(o) \tag{50}$$

§ 35 In the following chapters, and in almost all practical applications, we shall work only with the typical law of errors as our constant supposition This gives simplicity and clearness, and thus $a \pm b$ may be recommended as a short statement of the law of errors, $a = \lambda_1$ indicating a result of an observation found directly or indirectly by computation with observations, and $b = \sqrt{\lambda_2}$ expressing the mean error of the same result

By the *"weights"* of observations we understand numbers inversely proportional to the squares of the mean errors, consequently $v = \frac{1}{\lambda_2}$ The idea presents itself when we speak of the means of various numbers of observed values which have been obtained by the same method, as the latter numbers here, according to (37), represent the weights When v_r is the weight of the partial mean value m_r, the total mean value m must be computed according to the formula

$$m = \frac{m_1 v_1 + m_2 v_2 + \ldots + m_r v_r}{v_1 + v_2 + \ldots + v_r}, \tag{51}$$

which is analogous to the formula for the abscissa of the centre of gravity, if m_r is the abscissa of any single body, v_r its *weight* We speak also of the weights of single observations according to the above definition, and particularly in cases where we can only estimate the relative goodness of several observations in comparison to the trustworthiness of the means of various numbers of equally good observations

The phrase *probable error*, which we still find frequently employed by authors and observers, is for several reasons objectionable It can be used only with typical or at any rate symmetrical laws of errors, and indicates then the magnitude of errors for which the probabilities of smaller and larger errors are both equal to $\frac{1}{2}$ The simultaneous use of the ideas "mean error" and "probable error" causes confusion, and it is evidently the latter that must be abandoned, as it is less commonly applicable, and as it can only be computed in the cases of the typical law of errors by the previously computed mean error as $0\,6745\sqrt{\lambda_2}$, while on the other hand the computation of the mean error is quite independent of that of the probable error. As errors which are larger than the probable one, still frequently occur, this idea is not so well adapted as the mean error to serve as a limit between the frequent "small" errors and the rarer "large" ones The use of the probable error tempts us constantly to overvalue the degree of accuracy we have attained

More dangerous still is another confusion which now and then occurs, when the very expression mean error is used in the sense of the average error of the observed values according to their numerical values without regard to the signs This gives no sense, except when we are certain of a law of typical errors, and with such a one this „mean

error' is $\sqrt{\frac{2}{\pi}}\lambda_2$. The only reason which may be advanced in defence of the use of this idea is that we are spared some little computations, viz. some squarings and the extraction of a square root, which, however, we rarely need work out with more than three significant figures.

IX. FREE FUNCTIONS.

§ 30. The foregoing propositions concerning the laws of errors of functions — especially of linear functions — form the basis of the theory of computation with observed values, a theory which in several important things differs from exact mathematics. The result, particularly, is not an exact quantity, but always a law of errors which can be represented by its mean value and its mean error, just like the single observation. Moreover, the computation must be founded on a correct apprehension of what observations we may consider mutually unbound, another thing which is quite foreign to exact mathematics. For it is only upon the supposition that the result $R = r_1 o_1 + \ldots + r_n o_n = [ro]$ — observe the abbreviated notation — is a linear function of unbound observations only, $o_1, \ldots, o_n$, that we have demonstrated the rules of computation (35)

$$\lambda_1(R) = r_1\lambda_1(o_1) + \ldots + r_n\lambda_1(o_n) = [r\lambda_1(o)] \tag{52}$$

$$\lambda_2(R) = r_1^2\lambda_2(o_1) + \ldots + r_n^2\lambda_2(o_n) = [r^2\lambda_2(o)] \tag{53}$$

While the results of computations with observed quantities, taken singly, have laws of errors in the same way as the observations, they also resemble the observations in the circumstances that there can be bonds between them, and, unfortunately, there can be bonds between "results", even though they are derived from unbound observations. If only some observations have been employed in the computation of both $R' = [r'o]$ and $R'' = [r''o]$, these results will generally be bound to each other. This, however, does not prevent us from computing a law of errors, for instance for $aR' + bR''$. We can, at any rate, represent the function of the results directly as a function of the unbound observations, $o_1, \ldots, o_n$,

$$aR' + bR'' = [(ar' + br'')o] \tag{54}$$

This possibility is of some importance for the treatment of those cases in which the single observations are bound. They must be treated then just like results, and we must try to represent them as functions of the circumstances which they have in common, and which must be given instead of them as original observations. This may be difficult to do, but as a principle it must be possible and functions of bound observations must therefore always have laws of errors as well as others; only, in general, it is not possible to compute these laws of errors correctly simply by means of the laws of errors of the

observat. only, just as we cannot, in general, compute the law of errors for $aR'+bR''$ by means of the laws of errors for R' and R''.

In example 5, § 20, we found the mean error in the determination of a direction R between two points, which were given by bond-free and equally good ($\lambda_2(x) = \lambda_2(y) = 1$) measurements of their rectangular co-ordinates, viz. $\lambda_2(R) = \frac{2}{\Delta^2}$, and then, in example 6, we determined the angle V in a triangle whose points were determined in the same way. It seems an obvious conclusion then that, as $V = R'-R''$, we must have $\lambda_2(V) = \lambda_2(R')+\lambda_2(R'') = \frac{2}{\Delta'^2}+\frac{2}{\Delta''^2}$. But this is not correct, the solution is $\lambda_2(V) = \frac{\Delta^2+\Delta'^2+\Delta''^2}{\Delta'^2\Delta''^2}$, where Δ, Δ', and Δ'' are the sides of the triangle. The cause of this is, of course, that the co-ordinates of the angular point enter into both directions and bind R' and R'' together. But it is remarkable then that, when V is a right angle, the solutions are identical.

With equally good unbound observations, o_0, o_1, o_2, and o_3, we get

$$\lambda_2(o_2-2o_1+o_0) = 6\lambda_2(o)$$
$$\lambda_2(o_3-2o_2+o_1) = 6\lambda_2(o),$$

but

$$\lambda_2(o_3-3o_2+3o_1-o_0) = 20\lambda_2(o),$$

although $o_3-3o_2+3o_1-o_0 = (o_3-2o_2+o_1)-(o_2-2o_1+o_0)$, according to which we should expect to find

$$\lambda_2(o_3-3o_2+3o_1-o_0) = \lambda_2(o_3-2o_2+o_1)+\lambda_2(o_2-2o_1+o_0) = 12\lambda_2(o).$$

But if, on the other hand, we combine the two functions

$$R' = o_0+6o_1-4o_2 \quad \text{and} \quad R'' = 2o_1+3o_2-o_3,$$

where $\lambda_2(R') = 53\lambda_2(o)$ and $\lambda_2(R'') = 14\lambda_2(o)$, and from this compute λ_2 for any function $aR'+bR''$, then, curiously enough, we get as the correct result $\lambda_2(aR'+bR'') = (53a^2+14b^2)\lambda_2(o) = a^2\lambda_2(R')+b^2\lambda_2(R'')$.

Gauss's general prohibition against regarding results of computations — especially those of mean errors — from the same observations as analogous to unbound observations, has long hampered the development of the theory of observations.

To Oppermann and, somewhat later, to Helmert is due the honour of having discovered that the prohibition is not absolute, but that wide exceptions enable us to simplify our calculations. We must therefore study thoroughly the conditions on which actually existing bonds may be harmless.

Let $o_1, \ldots, o_n$ be mutually unbound observations with known laws of errors, $\lambda_1(o_i)$, $\lambda_2(o_i)$, of typical form. Let two general, linear functions of them be

$$[po] = p_1o_1+\cdots+p_no_n$$
$$[qo] = q_1o_1+\cdots+q_no_n$$

For these then we know the laws of errors

$$\left.\begin{array}{l}\lambda_1[po] = [p\lambda_1(o)],\quad \lambda_2[po] = [p^2\lambda_2(o)],\quad \lambda_r[po] = 0\\ \lambda_1[qo] = [q\lambda_1(o)],\quad \lambda_2[qo] = [q^2\lambda_2(o)],\quad \lambda_r[qo] = 0\end{array}\right\}\text{ for } r > 2$$

For a general function of these, $F = a[po] + b[qo]$, the correct computation of the law of errors by means of $F = [(ap + bq)o]$ will further give

$$\left.\begin{array}{rl}\lambda_1(F) &= (ap_1 + bq_1)\lambda_1(o_1) + \ldots + (ap_n + bq_n)\lambda_1(o_n) = \\ &= a\lambda_1[po] + b\lambda_1[qo]\end{array}\right\} \tag{55}$$

$$\left.\begin{array}{rl}\lambda_2(F) &= (ap_1 + bq_1)^2\lambda_2(o_1) + \ldots + (ap_n + bq_n)^2\lambda_2(o_n) = \\ &= a^2\lambda_2[po] + b^2\lambda_2[qo] + 2ab[pq\lambda_2(o)]\end{array}\right\} \tag{56}$$

$$\lambda_r(F) = 0 \text{ for } r > 2$$

It appears then, both that the mean values can be computed unconditionally, as if $[po]$ and $[qo]$ were unbound observations, and that the law of errors remains typical. Only in the square of the mean error there is a difference, as the term containing the factor $2ab$ in $\lambda_2(F)$ ought not to be found in the formula, if $[po]$ and $[qo]$ were not bound to one another

When consequently

$$[pq\lambda_2(o)] = p_1q_1\lambda_2(o_1) + \ldots + p_nq_n\lambda_2(o_n) = 0 \tag{57}$$

the functions $[po]$ and $[qo]$ can indeed be treated in all respects like unbound observations, for the law of errors for every linear function of them is found correctly determined also upon this supposition We call such functions mutually "free functions", and for such, consequently, the formula for the mean error

$$\lambda_2([po]a + [qo]b) = a^2[p^2\lambda_2(o)] + b^2[q^2\lambda_2(o)] \tag{58}$$

holds good

If this formula holds good for one set of finite values of a and b, it holds good for all

If two functions are mutually free, each of them is said to be "free of the other", and inversely

Example 1 The sum and difference of two equally good, unbound observations are mutually free.

Example 2 When the co-ordinates of a point are observed with equal accuracy and without any bonds, any transformed rectangular co-ordinates for the same will be mutually free

Example 3 The sum or the mean value of equally good, unbound observations is free of every difference between two of these, and generally also free of every (linear) function of such differences

Example 4 The differences between one observation and two other arbitrary, unbound observations cannot be mutually free

Example 5 Linear functions of unbound observations, which are all different, are always free

Example 6, Functions with a constant proportion cannot be mutually free

§ 37 In accordance with what we have now seen of free functions, corresponding propositions must hold good also of observations which are influenced by the same circumstances it is not necessary to respect all connecting bonds, It is possible that actually bound observations may be regarded as free The conditions on which this may be the case, must be sought, as in (67), by means of the mean errors caused by each circumstance and the coefficients by which the circumstance influences the several observations — Note particularly

If two observations are supposed to be connected by one single circumstance which they have in common, such a bond must not be left out of consideration, but is to be respected Likewise, if there are several bonds, each of which influences both observations in the same direction

If, on the other hand, some common circumstances influence the observations in the same direction, others in opposite directions, and if, moreover, one class must be supposed to work as forcibly as the other, the observations may possibly be free, and the danger of treating them as unbound is at any rate less than in the other cases

§ 38. Assuming that the functions of which we shall speak in the following are linear, or at any rate may be regarded as linear when expanded by Taylor's formula, because the errors are so small that we may reject squares and products of the deviations of the observations from fixed values, and assuming that the observations $o_1, \ldots o_n$, on which all the functions depend, are unbound, and that the values of $\lambda_2(o_1) \ldots \lambda_2(o_n)$ are given, we can now demonstrate a series of important propositions

Out of the total system of all functions

$$[po] = p_1 o_1 + \ldots + p_n o_n$$

of the given n observations we can arbitrarily select partial systems of functions, each partial system containing all those, which can be represented as functions of a number of $m < n$ mutually independent functions, representative of the system,

$$[ao] = a_1 o_1 + \ldots + a_n o_n$$

$$\ldots$$

$$[do] = d_1 o_1 + \ldots + d_n o_n,$$

of which no one can be expressed as a function of the others We can then demonstrate the existence of other functions which are free of every function belonging to the partial

system represented by $[ao], \dots, [do]$. It is sufficient to prove that such a function $[go] = g_1 o_1 + \dots + g_n o_n$ is free of $[ao], \dots, [do]$ in consequence of the equations $[ga\lambda_1] = 0, \dots, [gd\lambda_1] = 0$. For if so, $[go]$ must be free of every function of the partial system,

$$[(xa + \dots + zd)o] = x[ao] + \dots + z[do],$$

because

$$[g(xa + \dots + zd)\lambda_1] = x[ga\lambda_1] + \dots + z[gd\lambda_1] = 0.$$

Any function of the total system $[po]$ can now in one single way be resolved into a sum of two functions of the same observations, one of which is free of the partial system represented by $[ao] \dots [do]$, while the other belongs to this system.

If we call the free addendum $[p'o]$, this proposition may be written

$$[po] = [p'o] + \{x[ao] + \dots + z[do]\} \tag{59}$$

By means of the conditions of freedom, $[p'a\lambda_1] = \dots = [p'd\lambda_1] = 0$, all that concerns the unknown function $[p'o]$ can be eliminated. We find

$$\left.\begin{array}{l} [pa\lambda_1] = x[aa\lambda_1] + \dots + z[da\lambda_1] \\ \dots \\ [pd\lambda_1] = x[ad\lambda_1] + \dots + z[dd\lambda_1], \end{array}\right\} \tag{60}$$

from which we determine the coefficients $x \dots z$ unambiguously. The number m of these equations is equal to the number of the unknown quantities, and they must be sufficient for the determination of the latter, because, according to a well known proposition from the theory of determinants, the determinant of the coefficients

$$\begin{vmatrix} [aa\lambda_1], & \dots & [da\lambda_1] \\ \dots & & \dots \\ [ad\lambda_1], & \dots & [dd\lambda_1] \end{vmatrix} = \sum \begin{vmatrix} a_r, & \dots & a_s \\ \dots & & \dots \\ d_r, & \dots & d_s \end{vmatrix}^2 \lambda_1(o_r) \dots \lambda_1(o_s)$$

is positive, being a sum of squares, and cannot be $= 0$, unless at least one of the functions $[ao] \dots [do]$ could, contrary to our supposition, be represented as a function of the others.

From the values of $x \dots z$ thus found, we find likewise

$$[p'o] = [po] - x[ao] - \dots - z[do] \tag{61}$$

If $[po]$ belongs to the partial system represented by $[ao] \dots [do]$, the determination of $x \dots z$ expresses its coefficients in that system only, and then we get identically $[p'o] = 0$.

But if we take $[po]$ out of the partial system, then (61) gives us $[p'o]$ as different from zero and free of that partial system If $[po]-[qo]$ belongs to the partial system of $[ao]$ $[do]$, $[qo]$ must produce in this manner the very same free function as $[po]$.

Let $[po]$ $[ro]$ be $n-m$ functions, independent of one another and of the m functions $[ao]$ $[do]$, If we then find $[p'o]$ out of $[po]$ and $[r'o]$ out of $[ro]$ as the free functional parts in respect to $[ao]$ $[do]$, the n functions $[ao]$ $[do]$ and $[p'o]$ $[r'o]$ may be the representative functions of the total system of the functions of o_1 o_n, because no relation $a[p'o]+ \cdot + \delta[r'o] = 0$ is possible, for by (61) it might result in a relation $a[po]+ \quad + \delta[ro] + \varepsilon[ao] + \quad + \varphi[do] = 0$ in contradiction to the presumed representative character of $[po]$ $[ro]$ and $[ao]$, $[do]$

If we employ $[p'o]$ $[r'o]$ or other $n-m$ mutually independent functions

$$[go], \;. [ko],$$

all free of the partial set $[ao]$ $[do]$, as representative functions of another partial system of o_1 o_n, then every function of this system must be free of every function of the partial system $[ao]$. $[do]$ (Compare the introduction to this §) No other function of o_1 o_n can be free of $[ao]$, $[do]$ than those belonging to the system $[go]$ $[ko]$, otherwise we should have more than n independent functions of the n variables o_1 o_n

Thus selecting arbitrarily a partial system of functions of the observations o_1 , o_n we can — with reference to given squares of mean errors $\lambda_2(o_1)$ $\lambda_2(o_n)$ — distribute the linear homogeneous functions of these observations into three divisions:

1) the given partial system $[ao]$, $[do]$,
2) the partial system of functions $[go]$ $[ko]$, which are free of the former, and
3) all the rest, of which it is proved that every such function is always in only one way compounded by addition of one function of the first partial system to one of the second

The freedom of functions is a reciprocal property If the second partial system $[go]$ $[ko]$ were selected arbitrarily instead of the first $[ao]$ $[do]$, then only this latter would be found as the free functions in 2), the composition of every function in 3) would remain the same.

Example. Determine the parts of o_1+o_2, o_2+o_3, o_1+o_4, and o_2+o_3, which are free of o_1+o_2 and o_3+o_4, on the supposition that all 4 observations are equally exact and unbound

Answer: $\frac{1}{2}(o_1+o_2-o_3-o_4)$, etc.

§ 39. Like all other functions of the observations $o_1 \ldots o_n$, each of these observed values, for instance o_i, is the sum of two quantities, one o_i' belonging to the system of $[ao] \ldots [do]$, the other o_i'' to the partial system of $[go] \ldots [ko]$, which is free of this. But from $o_i = o_i'' + o_i'$ follows, generally, that $[po] = [po'] + [po'']$, and $[po'']$ evidently belongs to the system of $[ao] \ldots [do]$, $[po']$ to the system which is free of this. Accordingly there must between the n functions $o_1' \ldots o_n'$ exist m relations $[ao'] = \ldots = [do'] = 0$, likewise $n - m$ relations $[go''] = \ldots = [ko''] = 0$, between $o_1'' \ldots o_n''$.

§ 40. That the functions of observations can be split up, in an analogous way, into three or more free quantities, is of no consequence for the following, *except when we imagine this operation to be carried through to the utmost*. It is easy enough to see, however, that also the partial systems of functions can be split up. We could, for instance, among the representatives $[ao] \ldots [do]$ of one partial system select a smaller number $[ao] \ldots [bo]$, and from the others $[co] \ldots [do]$, according to (37), separate the functions $[c'o] \ldots [d'o]$ which were free of $[ao] \ldots [bo]$. $[c'o] \ldots [d'o]$ would then represent the subsystem of functions, free of $[ao] \ldots [bo]$, within the partial system $[ao] \ldots [do]$, and in this way we may continue till all representative functions are mutually free, every single one of all the rest. Such a collection of representative functions we call a *complete set of free functions*. Their number is sufficient to enable us to express all the observations, and all functions of these observations, as functions of them, and their mutual freedom has the effect that they can be treated, by all computations of laws of errors, quite like unbound observations, and thus wholly replace the original observations.

§ 41. The mathematical theory of the transformations of observations into free functions is analogous to the theory of the transformation of rectangular co-ordinates (comp. § 36, example 2), and is treated in several text-books of the higher algebra and determinants under the name of the theory of the orthogonal substitutions. I shall here enter into those propositions only, which we are to use in what follows.

When we have transformed the unbound observations $o_1 \ldots o_n$ into the complete set of free functions $[ao]$, $[b'o] \ldots [d^{\nu}o]$, it is often important to be able to undertake the opposite transformation back to the observations. This is very easily done, for we have

$$o_i = \left\{ \frac{a_i}{[aa\lambda_2]} [ao] + \ldots + \frac{d_i^{\nu}}{[d^{\nu} d^{\nu} \lambda_2]} [d^{\nu} o] \right\} \lambda_2(o_i), \tag{62}$$

which is demonstrated by substitution in the equations for the direct transformation

$$[ao] = a_1 o_1 + \ldots + a_n o_n$$

$$\ldots$$

$$[d^{\nu} o] = d_1^{\nu} o_1 + \ldots + d_n^{\nu} o_n,$$

because $[ab'\lambda_2] = [ad^{\nu}\lambda_2] = \ldots = [b'd^{\nu}\lambda_2] = 0$.

As the original observations, considered as functions of the transformed observations $[ao]$ ... $[d^{\nu}o]$, must be mutually free, just as well as the latter are free functions of the former, we find by computing the squares of the mean errors $\lambda_1(o_i)$ and the equation that expresses the formal condition that o_i is free of o_k, two of the most remarkable properties of the orthogonal substitutions:

$$\frac{1}{\lambda_1(o_i)} = \frac{a_i^2}{[aa\lambda_1]} + \dots + \frac{d_i^{\nu 2}}{[d^{\nu}d^{\nu}\lambda_1]} \tag{63}$$

and

$$0 = \frac{a_i a_k}{[aa\lambda_1]} + \dots + \frac{d_i^{\nu} d_k^{\nu}}{[d^{\nu}d^{\nu}\lambda_1]} \tag{64}$$

If all observations and functions are stated with their respective mean error as unity, or are divided by their mean error, a reduction which gives also a more elegant form to all the preceding equations, the sum of the squares of the thus reduced observations is not changed by any (orthogonal) transformation into a complete set of free functions.

We have

$$\frac{[ao]^2}{\lambda_1[ao]} + \dots + \frac{[d^{\nu}o]^2}{\lambda_1[d^{\nu}o]} = \frac{o_1^2}{\lambda_1(o_1)} + \dots + \frac{o_n^2}{\lambda_1(o_n)}, \tag{65}$$

which, pursuant to the equations (63) and (64), is easily demonstrated by working out the sums of the squares in the numerators on the left side of the equation. As this equation is identical, the same proposition holds good also, for instance, of the differences between o_1 ... o_n and n arbitrarily selected variables corresponding to them v_1 ... v_n, and of the corresponding differences between the values of the functions. Also here is

$$\left.\begin{aligned}\frac{([ao]-[av])^2}{\lambda_1[ao]} + \dots + \frac{([d^{\nu}o]-[d^{\nu}v])^2}{\lambda_1[d^{\nu}o]} = \frac{[a(o-v)]^2}{[aa\lambda_1]} + \dots + \frac{[d^{\nu}(o-v)]^2}{[d^{\nu}d^{\nu}\lambda_1]} = \\ = \frac{(o_1-v_1)^2}{\lambda_1(o_1)} + \dots + \frac{(o_n-v_n)^2}{\lambda_1(o_n)}.\end{aligned}\right\} \tag{66}$$

§ 42. For the practical computation of a complete set of free functions it will be the easiest way to bring forward the functions of such a set one by one. In this case we must select a sufficient number of functions and fix the order in which these are to be taken into consideration. For a moment we can imagine this order to be arbitrary.

The function $[ao]$, which is the first in this list, is now, unchanged, taken into the transformed set. By multiplying the selected function by suitable constants of the form $\frac{[ba\lambda_1]}{[aa\lambda_1]}$, and subtracting the products from the remaining functions $[bo]$ in the list, we can, according to § 38, from each of these separate the addendum which is free of the selected function. Of these then the one which is founded on function Nr 2 on the list is taken into the transformed set. This function is multiplied in the same way and subtracted from

the still remaining functions, so that they give up the addenda which are free of both the selected functions, and so on. The following schedule shows the course of the operation for the case $n = 4$.

Functions	Coefficients $\lambda_1(o_1)\ \lambda_1(o_2)\ \lambda_1(o_3)\ \lambda_1(o_4)$	Sums of the Products	Rule of Computation
$[ao]$	$a_1\ a_2\ a_3\ a_4$	$[aa\lambda]\ [ab\lambda]\ [ac\lambda]\ [ad\lambda]$	$[ao]$ is selected
$[bo]$	$b_1\ b_2\ b_3\ b_4$	$[ba\lambda]\ [bb\lambda]\ [bc\lambda]\ [bd\lambda]$	$[bo]-[ao]\ [ba\lambda]:[aa\lambda]=[b'o]$
$[co]$	$c_1\ c_2\ c_3\ c_4$	$[ca\lambda]\ [cb\lambda]\ [cc\lambda]\ [cd\lambda]$	$[co]-[ao]\ [ca\lambda]:[aa\lambda]=[c'o]$
$[do]$	$d_1\ d_2\ d_3\ d_4$	$[da\lambda]\ [db\lambda]\ [dc\lambda]\ [dd\lambda]$	$[do]-[ao]\ [da\lambda]:[aa\lambda]=[d'o]$
$[b'o]$	$b'_1\ b'_2\ b'_3\ b'_4$	$[b'b'\lambda]\ [b'c'\lambda]\ [b'd'\lambda]$	$[b'o]$ is selected, is free of $[ao]$
$[c'o]$	$c'_1\ c'_2\ c'_3\ c'_4$	$[c'b'\lambda]\ [c'c'\lambda]\ [c'd'\lambda]$	$[c'o]-[b'o]\ [c'b'\lambda]:[b'b'\lambda]=[c''o]$
$[d'o]$	$d'_1\ d'_2\ d'_3\ d'_4$	$[d'b'\lambda]\ [d'c'\lambda]\ [d'd'\lambda]$	$[d'o]-[b'o]\ [d'b'\lambda]:[b'b'\lambda]=[d''o]$
$[c''o]$	$c''_1\ c''_2\ c''_3\ c''_4$	$[c''c''\lambda]\ [c''d''\lambda]$	$[c''o]$ is selected, is free of $[b'o]$ and $[ao]$
$[d''o]$	$d''_1\ d''_2\ d''_3\ d''_4$	$[d''c''\lambda]\ [d''d''\lambda]$	$[d''o]-[c''o]\ [d''c''\lambda]:[c''c''\lambda]=[d'''o]$
$[d'''o]$	$d'''_1\ d'''_2\ d'''_3\ d'''_4$	$[d'''d'''\lambda]$	$[d'''o]$ is free of $[c''o]$, $[b'o]$, and $[ao]$

The computations of the sums of the products (in which for the sake of brevity we have written λ for $\lambda_1(o)$) could be made all through by means of the single coefficients in the transformed functions, as it must be done in the beginning by means of the coefficients in the original functions. It is much easier, however, (particularly if for some reason or other we might otherwise do without the computation of the coefficients of the transformed functions), to make use, for this purpose, of the following remarkable property of these sums of the products. We have, for instance,

$$\left.\begin{aligned}
[b'c'\lambda] &= \left[\left(b-a\frac{[ba\lambda]}{[aa\lambda]}\right)\left(c-a\frac{[ca\lambda]}{[aa\lambda]}\right)\lambda\right] = \\
&= [bc\lambda]-[ac\lambda]\frac{[ba\lambda]}{[aa\lambda]}-[ba\lambda]\frac{[ca\lambda]}{[aa\lambda]}+[aa\lambda]\frac{[ba\lambda]}{[aa\lambda]}\,\frac{[ca\lambda]}{[aa\lambda]} = \\
&= [bc\lambda]-[ac\lambda]\ [ba\lambda]:[aa\lambda]
\end{aligned}\right\} \qquad (67)$$

Consequently, the same general rule of computation as, according to the schedule, holds good of the functions and their coefficients, holds good also of the sums of the products and of the squares. The schedule gets the following appendix:

$$[bb\lambda]-[ab\lambda]\cdot[ba\lambda]:[aa\lambda]=[b'b'\lambda],\quad [bc\lambda]-[ac\lambda]\cdot[ba\lambda]:[aa\lambda]=[b'c'\lambda],\quad [bd\lambda]-[ad\lambda]\cdot[ba\lambda]:[aa\lambda]=[b'd'\lambda]$$
$$[cb\lambda]-[ab\lambda]\cdot[ca\lambda]:[aa\lambda]=[c'b'\lambda],\quad [cc\lambda]-[ac\lambda]\cdot[ca\lambda]:[aa\lambda]=[c'c'\lambda],\quad [cd\lambda]-[ad\lambda]\cdot[ca\lambda]:[aa\lambda]=[c'd'\lambda]$$
$$[db\lambda]-[ab\lambda]\cdot[da\lambda]:[aa\lambda]=[d'b'\lambda],\quad [dc\lambda]-[ac\lambda]\cdot[da\lambda]:[aa\lambda]=[d'c'\lambda],\quad [dd\lambda]-[ad\lambda]\cdot[da\lambda]:[aa\lambda]=[d'd'\lambda]$$

$$[c'c'\lambda]-[b'c'\lambda]\cdot[c'b'\lambda]:[b'b'\lambda]=[c''c''\lambda],\quad [c'd'\lambda]-[b'd'\lambda]\cdot[c'b'\lambda]:[b'b'\lambda]=[c''d''\lambda]$$
$$[d'c'\lambda]-[b'c'\lambda]\cdot[d'b'\lambda]:[b'b'\lambda]=[d''c''\lambda],\quad [d'd'\lambda]-[b'd'\lambda]\cdot[d'b'\lambda]:[b'b'\lambda]=[d''d''\lambda]$$

$$[d''d''\lambda]-[c''d''\lambda]\cdot[d''c''\lambda]:[c''c''\lambda]=[d'''d'''\lambda]$$

As will be seen, there is a check by means of double computation for each of the sums of the products properly so called. The sums of the squares are of special importance as they are the squares of the mean errors of the transformed functions, $\lambda_1[ao]=[aa\lambda]$, $\lambda_1[b'o]=[b'b'\lambda]$, $\lambda_1[c''o]=[c''c''\lambda]$, and $\lambda_1[d'''o]=[d'''d'''\lambda]$

Example. Five equally good, unbound observations o_1, o_2, o_3, o_4, and o_5 represent values of a table with equidistant arguments. The function tabulated is known to be an integral algebraic one, not exceeding the 3rd degree. The transformation into free functions is to be carried out, in such a way that the higher differences are selected before the lower ones (Because Δ^4, certainly, Δ^3 etc., possibly, represent equations of condition). With symbols for the differences, and with $\lambda_1(o_i)=1$, we have then.

Function	Coefficients					Sums of the Products					Factors
o_3	$0o_1$	$+0o_2$	$+1o_3$	$+0o_4$	$+0o_5$	1	−1	−2	3	6	$-\frac{6}{70}$
$V\Delta o_3$	0	0	−1	1	0		2	3	−6	−10	$\frac{1}{7}$
$\Delta^2 o_2$	0	1	−2	1	0		4	0	−10	−20	$\frac{2}{7}$
$V\Delta^3 o_2$	0	−1	3	−3	1	3	−6	10	20	35	$-\frac{1}{2}$
$\Delta^4 o_1$	1	−4	6	−4	1	6	−10	−20	35	70	is selected
$o_3-\frac{3}{35}\Delta^4 o_1$	$-\frac{3}{35}$	$\frac{12}{35}$	$\frac{17}{35}$	$\frac{12}{35}$	$-\frac{3}{35}$	$\frac{17}{35}$	$-\frac{1}{7}$	$-\frac{2}{7}$	0		0
$V\Delta o_3+\frac{1}{7}\Delta^4 o_1$	$\frac{1}{7}$	$-\frac{4}{7}$	$-\frac{1}{7}$	$\frac{3}{7}$	$\frac{1}{7}$	$-\frac{1}{7}$	$\frac{4}{7}$	$\frac{1}{7}$	−1		$\frac{2}{5}$
$\Delta^2 o_2+\frac{2}{7}\Delta^4 o_1$	$\frac{2}{7}$	$-\frac{1}{7}$	$-\frac{2}{7}$	$-\frac{1}{7}$	$\frac{2}{7}$	$-\frac{2}{7}$	$\frac{1}{7}$	$\frac{2}{7}$	0		0
$V\Delta^3 o_2-\frac{1}{2}\Delta^4 o_1$	$-\frac{1}{2}$	1	0	−1	$\frac{1}{2}$	0	−1	0	$\frac{5}{2}$		is selected
$o_3-\frac{3}{35}\Delta^4 o_1$	$-\frac{3}{35}$	$\frac{12}{35}$	$\frac{17}{35}$	$\frac{12}{35}$	$-\frac{3}{35}$	$\frac{17}{35}$	$-\frac{1}{7}$	$-\frac{2}{7}$			1
$V\Delta o_3+\frac{2}{5}V\Delta^3 o_2-\frac{2}{35}\Delta^4 o_1$	$-\frac{2}{35}$	$-\frac{6}{35}$	$-\frac{1}{7}$	$\frac{1}{35}$	$\frac{12}{35}$	$-\frac{1}{7}$	$\frac{6}{35}$	$\frac{1}{7}$			$-\frac{1}{2}$
$\Delta^2 o_2+\frac{2}{7}\Delta^4 o_1$	$\frac{2}{7}$	$-\frac{1}{7}$	$-\frac{2}{7}$	$-\frac{1}{7}$	$\frac{2}{7}$	$-\frac{2}{7}$	$\frac{1}{7}$	$\frac{2}{7}$			is selected
$o_3+\Delta^2 o_2+\frac{1}{5}\Delta^4 o_1$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0				are free
$V\Delta o_3-\frac{1}{2}\Delta^2 o_2+\frac{2}{5}V\Delta^3 o_2-\frac{1}{5}\Delta^4 o_1$	$-\frac{1}{5}$	$-\frac{1}{10}$	0	$\frac{1}{10}$	$\frac{1}{5}$	0	$\frac{1}{10}$				are both selec

The complete set of free observations and the squares of their mean errors are thus

$$
\begin{array}{llll}
(0) = & o_3+\Delta^2 o_3+\frac{1}{5}\Delta^4 o_3 & = \frac{1}{5}(o_1+o_2+o_3+o_4+o_5), & \lambda_2(0) = \frac{1}{5} \\
(1) = & V\Delta o_3-\frac{1}{6}\Delta^2 o_3+\frac{2}{5}(V\Delta^3 o_3-\frac{1}{6}\Delta^4 o_3) & = \frac{1}{10}(-2o_1-o_2+o_4+2o_5), & \lambda_2(1) = \frac{1}{10} \\
(2) = & \Delta^2 o_3+\frac{2}{7}\Delta^4 o_3 & = \frac{1}{7}(2o_1-o_2-2o_3-o_4+2o_5), & \lambda_2(2) = \frac{2}{7} \\
(3) = & V\Delta^3 o_3-\frac{1}{2}\Delta^4 o_3 & = \frac{1}{2}(-o_1+2o_2-2o_4+o_5), & \lambda_2(3) = \frac{5}{2} \\
(4) = & \Delta^4 o_3 & = o_1-4o_2+6o_3-4o_4+o_5, & \lambda_2(4) = 70
\end{array}
$$

Through this and the preceding chapter we have got a basis which will generally be sufficient for computations with observations and, in a wider sense, for computations with numerical values which are not given in exact form, but only by their laws of errors We can, in the first place, compute the law of errors for a given, linear function of reciprocally free observations whose laws of presumptive errors we know By this we can solve all problems in which there is not given a greater number of observations, and other more or less exact data, than of the reciprocally independent unknown values of the problem. When we, in such cases, by the means of the exact mathematics, have expressed each of the unknown numbers as a function of the given observations, and when we have succeeded in bringing these functions into a linear form, then we can, by (35), compute the laws of errors for each of the unknown numbers

Such a solution of a problem may be looked upon as a transformation, by which n observed or in other ways given values are transformed into n functions, each corresponding to its particular value among the independent, unknown values of the problem It lies often near thus to look upon the solution of a problem as a transformation, when the solution of the problem is not the end but only the means of determining other unknown quantities, perhaps many other, which are all explicit functions of the independent unknowns of the problem. Thus, for instance, we compute the 6 elements of the orbit of a planet by the rectascensions and declinations corresponding to 3 times, not precisely as our end, but in order thereby to be able to compute ephemerides of the future places of the planet. But while the validity of this view is absolute in exact mathematics, it is only limited when we want to determine the presumptive laws of errors of sought functions by the given laws of errors for the observations Only the mean values, sought as well as given, can be treated just as exact quantities, and with these the general linear transformation of n given into n sought numbers, with altogether n^2 arbitrary constants, remains valid, as also the employment of the found mean numbers as independent variables in the mean value of the explicit functions.

If we want also correctly to determine the mean errors, we may employ no other transformation than that into free functions And if, to some extent, we may choose the

independent unknowns of the problem as we please, we may often succeed in carrying through the treatment of a problem by transformation into free functions, for an unknown number may be chosen quite arbitrarily in all its n coefficients, and each of the following unknowns looses, as, a function of the observations, only an arbitrary coefficient in comparison to the preceding one, even the n^{th} unknown can still get an arbitrary factor Altogether are $\frac{1}{2}n(n+1)$ of the n^2 coefficients of these transformations arbitrary

But *if the problem does not admit of any solution through a transformation into* free functions, the mean errors for the several unknowns, no matter how many there may be, can be computed only in such a way that each of the sought numbers are directly expressed as a linear function of the observations The same holds good also when the laws of errors of the observations are not typical, and we are to examine how it is with λ_3 and the higher half-invariants in the laws of errors of the sought functions

Still greater importance, nay a privileged position as the only legitimate proceeding, gets the transformation into a complete set of free functions in the over-determined problems, which are rejected as self-contradictory in exact mathematics When we have a collection of observations whose number is greater than the number of the independent unknowns of the problem, then the question will be to determine laws of actual errors from the standpoint of the observations. We must mediate between the observations that contradict one another, in order to determine their mean numbers, and the discrepancies themselves must be employed to determine their mean deviations, etc But as we have not to do with repetitions, the discrepancies conceal themselves behind the changes of the circumstances and require transformations for their detection All the functions of the observations which, as the problem is over-determined, have theoretically necessary values, as, for instance, the sum of the angles of a plane triangle, must be selected for special use Besides, those of the unknowns of the problem, to the determination of which the theory does not contribute, must come forth by the transformation by which the problem is to be solved

As we shall see in the following chapters on Adjustment, it becomes of essential moment here that we transform into a system of free functions The transformation begins with mutually free observations, and must not itself introduce any bond, because the transformed functions in various ways must come forth as observations which determine laws of actual errors.

X. ADJUSTMENT.

§ 48. Pursuing the plan indicated in § 5 we now proceed to treat the determination of laws of errors in some of the cases of observations made under varying or different

essential circumstances But here we must be content with very small results The general problem will hardly ever be solved The necessary equations must be taken from the totality of the hypotheses or theories which express all the terms of each law of error — say their half-invariants — as functions of the varying or wholly different circumstances of the observations Without great regret, however, the multiplicity of these *theoretical equations* can be reduced considerably, if we suppose all the laws of errors to be exclusively of the typical form.

For each observation we need then only two theoretical equations, one representing its presumptive mean value $\lambda_1(o_i)$, the other the square of its mean error $\lambda_2(o_i)$, as functions of the essential circumstances But the theoretical equations will generally contain other unknown quantities, the arbitrary constants of the theory, and these must be eliminated or determined together with the laws of errors. The complexity is still great enough to require a further reduction

We must, preliminarily at all events, suppose the mean errors to be given directly by theory, or at least their mutual ratios, the weights If not, the problems require a solution by the indirect proceeding Hypothetical assumptions concerning the $\lambda_2(o_i)$ are used in the first approximation and checked and corrected by special operations which, as far as possible, we shall try to expose beside the several solutions, using for brevity the word "criticism" for those and other operations connected with them

But even if we confine our theoretical equations to the presumptive means $\lambda_1(o_i)$ and the arbitrary unknown quantities of the theory, the solutions will only be possible if we further suppose the theoretical equations to be linear or reducible to this form Moreover, it will generally be necessary to regard as exactly given many quantities really found by observation, on the supposition only that the corresponding mean errors will be small enough to render such irregularity inoffensive

In the solution of such problems we must rely on the found propositions about functions of observations with exactly given coefficients In the theoretical equations of each problem sets of such functions will present themselves, some functions appearing as given, others as required The observations, as independent variables of these functions, are, now the given observed values o_i, now the presumptive means $\lambda_1(o_i)$; the latter are, for instance, among the unknown quantities required for the exact satisfaction of the theoretical equations

What is said here provisionally about the problems that will be treated in the following, can be illustrated by the simplest case (discussed above) of n repetitions of the same observation, resulting in the observed values $o_1, \ldots, o_n$ If we here write the theoretical equations without introducing any unnecessary unknown quantities, they will show the forms $0 = \lambda_1(o_i) - \lambda_1(o_k)$ or, generally, $0 = \lambda_1[a(o_i - o_k)]$ But these equations are

evidently not sufficient for the determination of any $\lambda_1(o_i)$, which they only give if another $\lambda_1(o_k)$ is found beforehand. The sought common mean cannot be formed by the introduction of the observed values into any function $[u(o_i - o_k)]$, these erroneous values of the functions being useful only to check $\lambda_1(o_i)$ by our criticism. But we must remember what we know about free functions that the whole system of these functions $[u(o_i - o_k)]$ is only a partial system, with $n-1$ differences $o_i - o_k$ as representatives. The only n^{th} functions which can be free of this partial system, must evidently be proportional to the sum $o_1 + \ldots + o_n$, and by this we find the sought determination by

$$\lambda_1(o_i) = \frac{1}{n}(o_1 + \ldots + o_n),$$

the presumptive mean being equal to the actual mean of the observed values

If we thus consider a general series of unbound observations, $o_1, \ldots, o_n$, it is of the greatest importance to notice first that two sorts of special cases may occur, in which our problem may be solved immediately It may be that the theoretical equations concerning the observations leave some of the observations, for instance o_1, quite untouched; it may be also that the theory fully determines certain others of the observations, for instance o_n

In the former case, that is when none of all the theories in any way concern the observation o_1, it is evident that the observed value o_1 must be approved unconditionally Even though this observation does not represent any mean value found by repetitions, but stands quite isolated, it must be accepted as the mean $\lambda_1(o_1)$ in its law of presumptive errors, and the corresponding square of the mean error $\lambda_2(o_1)$ must then be taken, unchanged, from the assumed investigations of the method of observation

If, in the latter case, o_n is an observation which directly concerns a quantity that can be determined theoretically (for instance the sum of the angles of a rectilinear triangle), then it is, as such, quite superfluous as long as the theory is maintained, and then it must in all further computations be replaced by the theoretically given value; and in the same way $\lambda_2(o_n)$ must be replaced by zero, as the square of the mean error on the errorless theoretical value

The only possible meaning of such superfluous observations must be to test the correctness of the theory for approbation or rejection (a third result is impossible when we are dealing with any real theory or hypothesis), or to be used in the criticism

In such a test it must be assumed that the theoretical value corresponding to o_n, which we will call u_n, is identical with the mean value in the law of presumptive errors for o_n, consequently, that $u_n = \lambda_1(o_n)$, and the condition of an affirmative result must be obtained from the square of the deviation, $(o_n - u_n)^2$ in comparison with $\lambda_2(o_n)$. The

equation $(o_n - u_n)^2 = \lambda_2(o_n)$ need not be exactly satisfied, but the approximation must at any rate be so close that we may expect to find $\lambda_2(o_n)$ coming out as the mean of numerous observed values of $(o_n - u_n)^2$ Compare § 34

§ 44 If then all the observations $o_1 \ldots o_n$ fall under one or the other of these two cases, the matter is simple enough. But generally the observations o_i will be connected by theoretical equations of condition which, separately, are insufficient for the determination of the single ones Then the question is whether we can transform the series of observations in such a way that a clear separation between the two opposite relations to the theory can be made, so that some of the transformed functions of the observations, which must be mutually free in order to be treated as unbound observations, become quite independent of the theory, while the rest are entirely dependent on it This can be done, and the computation with observations in consequence of these principles, is what we mean by the word "adjustment"

For as every theory can be fully expressed by a certain number, $n - m$, of theoretical equations which give the exact values of the same number of mutually independent linear functions, and as we are able, as we have seen, from every observation or linear function of the observations, in one single way, to separate a function which is free and independent of these just named theoretically given functions, and which must thus enter into another system, represented by m functions, this system must include all those functions of the observations which are independent of the theory and cannot be determined by it Each of the thus mutually separated systems can be imagined to be represented, the theoretical system by $n-m$, the non-theoretical or empirical system by m mutually free functions, which together represent all observations and all linear functions of the same, and which may be looked upon as a complete, transformed system of free functions, consequently as unbound observations The two systems can be separated in a single way only, although the representation of each partial system, by free functions, can occur in many ways

It is *the idea of the adjustment*, by means of this transformation, to give the theory its due and the observations theirs, in such a way that every function of the theoretical system, and particularly the $n - m$ free representatives of the same, are exchanged, each with its theoretically given value, which, pursuant to the theory, is free of error On the other hand, every function of the empiric system and, particularly, its m free representatives remain unchanged as the observations determine them Every general function of the n observations $[do]$ and, particularly, the observations themselves are during the adjustment split into two univocally determined addenda the theoretical function $[d'o]$, which should have a fixed value D', and the non-theoretical one $[d''o]$ The former $[d'o]$ is by the adjustment changed into D' and made errorless, the latter is not changed at all The result of the adjustment, $D' + [d''o]$, is called the adjusted value of the function, and may

be indicated as $[du]$, the adjusted values of the observations themselves being written $u_1 \ldots u_n$. The forms of the functions are not broken, as the distributive principle $f(x+y) = f(x)+f(y)$ holds good of every homogeneous linear function.

The determination of the adjusted values is analogous to the formation of the mean values of laws of errors by repetitions. For theoretically determined functions the adjusted value is the mean value on the very law of presumptive errors, for the functions that are free of the whole theory, we have the extreme opposite limiting case, mean values represented by an isolated, single observation. In general the adjusted values $[du]$ are analogous to actual mean values by a more or less numerous series of repetitions. For while $\lambda_2([do]) = \lambda_2[d'o]+\lambda_2[d''o]$, we have $\lambda_2[du] = \lambda_2(D')+\lambda_2[d''u] = \lambda_2[d''o]$, consequently smaller than $\lambda_2[do]$. The ratio $\frac{\lambda_2[do]}{\lambda_2[du]}$ is analogous to the number of the repetitions or the weight of the mean value.

§ 45. By "*criticism*" we mean the trial of the — hypothetical or theoretical — suppositions, which have been made in the adjustment, with respect to the mean errors of the observations, new determinations of the mean errors, analogous to the determinations by the square of the mean deviations, μ_2, will, eventually also fall under this. The basis of the criticism must be taken from a comparison of the observed and the adjusted values, for instance the differences $[do]-[du]$. According to the principle of § 34 we must expect the square of such a difference, on an average, to agree with the square of the corresponding mean error, $\lambda_2([do]-[du])$, but as $[do]-[du] = [d'o]-D'$, and $\lambda_2[d'o] = \lambda_2[do]-\lambda_2[du]$, we get

$$\lambda_2([do]-[du]) = \lambda_2[do]-\lambda_2[du], \tag{68}$$

which, by way of parenthesis, shows that the observed and the adjusted values of the same function or observation cannot in general be mutually free. We ought then to have

$$\frac{([do]-[du])^2}{\lambda_2[do]-\lambda_2[du]} = 1 \tag{69}$$

on the average, and for a sum of terms of this form we must expect the mean to approach the number of the terms, nota bene, if there are no bonds between the functions $[do]-[du]$; but in general such bonds will be present, produced by the adjustment or by the selection of the functions.

It is no help if we select the original and unbound observations themselves, and consequently form sums such as

$$\left[\frac{(o-u)^2}{\lambda_2(o)-\lambda_2(u)}\right],$$

for after the adjustment and its change of the mean errors, $u_1 \ldots u_n$ are not generally free functions such as $o_1 \ldots o_n$. Only one single choice is immediately safe, viz., to stick to the system of the mutually free functions which, in the adjustment, have themselves

represented the observations the $n-m$ theoretically given functions and the m which the adjustment determines by the observations. Only of these we know that they are free both before and after the adjustment. And as the differences of the last-mentioned m functions identically vanish, the criticism must be based upon the $n-m$ terms corresponding to the theoretically free functions $[ao] = A$, $[b'o] = B'$ of the series

$$\frac{([ao]-A)^2}{\lambda_1[ao]-\lambda_1[au]} + \cdot + \frac{([b'o]-B')^2}{\lambda_1[b'o]-\lambda_1[b'u]} = \frac{([ao]-A)^2}{[aa\lambda_2]} + \quad + \frac{([b'o]-B')^2}{[b'b'\lambda_2]}, \quad (70)$$

the sum of which must be expected to be $= n-m$.

Of course we must not expect this equation to be strictly satisfied, according to the second equation (46) the square of the mean error on 1, as the expected value of each term of the series, ought to be put down $= 2$, for the whole series, consequently, we can put down the expected value as $n-m \pm \sqrt{2(n-m)}$.

But now we can make use of the proposition (66) concerning the free functions. It offers us the advantage that we can base the criticism on the deviations of the several observations from their adjusted values, the latter, we know, being such a special set of values as may be compared to the observations like $v_1, \ldots, v_n$ loc. cit., $u_1, \ldots u_n$ are only distinguished from $v_1 \ldots v_n$ by giving the functions which are free of the theory the same values as the observations. We have consequently

$$\frac{([ao]-A)^2}{[aa\lambda_2]} + \quad + \frac{([b'o]-B')^2}{[b'b'\lambda_2]} = \left[\frac{(o-u)^2}{\lambda_1(o)}\right] = n-m \pm \sqrt{n-m} \quad (71)$$

If we compare the sum on the right side in this expression with the above mentioned $\left[\frac{(o-u)^2}{\lambda_1(o)-\lambda_1(u)}\right]$, which we dare not approve on account of the bonds produced by the adjustment, then there is no decided contradiction between putting down $\left[\frac{(o-u)^2}{\lambda_1(o)}\right]$ at the smaller value $n-m$ only, while $\left[\frac{(o-u)^2}{\lambda_1(o)-\lambda_1(u)}\right]$, by the diminution of the denominators, can get the value n, only we can get no certainty for it.

The ratios between the corresponding terms in these two sums of squares, consequently $\frac{\lambda_1(o)-\lambda_1(u)}{\lambda_1(o)} = 1 - \frac{\lambda_1(u)}{\lambda_1(o)}$, we call "scales", viz *scales* for measuring the influence of the adjustment on the single observation. More generally we call

$$1-\frac{\lambda_1[du]}{\lambda_1[do]} \text{ the scale for the function } [do] \quad (72)$$

If the scale for a function or observation has its greatest possible value, viz 1, $\lambda_1[du] = 0$. The theory has then entirely decided the result of the adjustment. But if the scale sinks to its lowest limit $= 0$, we get just the reverse $\lambda_1[du] = \lambda_1[do]$, i. e. the theory has had no influence at all; the whole determination is based on the accidental

value of the observation, and for observations in this case we get $\frac{(o-u)^2}{\lambda_1(o)-\lambda_2(u)} = \frac{0}{0}$ Even though the scale has a finite, but very small value it will be inadmissible to depend on the value of such a term becoming = 1 We understand now, therefore, the superiority of the sum of the squares $\left[\frac{(o-u)^2}{\lambda_1(o)}\right] = n - m$ to the sum of the squares $\left[\frac{(o-u)^2}{\lambda_1(o)-\lambda_2(u)}\right] = n$ as a bearer of the summary criticism

We may also very well, on principle, sharpen the demand for adjustment on the part of the criticism, so that not only the whole sum of the squares $\left[\frac{(o-u)^2}{\lambda_1(o)}\right]$ must approach the value $n-m$, but also partial sums, extracted from the same, or even its several terms, must approach certain values Only, they are not to be added up as numbers of units, but must be sums of the scales of the corresponding terms So much we may trust to the sum of the squares $\left[\frac{(o-u)^2}{\lambda_1(o)-\lambda_2(u)}\right]$ that this principle, when judiciously applied, may be considered as fully justified

The sum of the squares $\left[\frac{(o-u)^2}{\lambda_1(o)}\right]$ possesses an interesting property which all other authors have used as the basis of the adjustment, under the name of *"the method of the least squares"* The above sum of the squares gets by the adjustment the least possible value that $\left[\frac{(o-v)^2}{\lambda_1(o)}\right]$ can get for values $v_1 \ldots v_n$ which satisfy the conditions of the theory. The proposition (66) concerning the free functions shows that the condition of this minimum is that $[d'o] = [d'u]$, $[d''o] = [d''u]$ for all the free functions which are determined by the observations, consequently just by putting for each o the corresponding adjusted value u

§ 46 The carrying out of adjustments depends of course to a high degree on the form in which the theory is given The theoretical equations will generally include some observations and, beside these, some unknown quantities, elements, in smaller number than those of the equations, which we just want to determine through the adjustment This general form, however, is unpractical, and may also easily be transformed through the usual mathematical processes of elimination We always go back to one or the other of two extreme forms which it is easy to handle *either*, we assume that all the elements are eliminated, so that the theory is given as above assumed by $n-m$ linear equations of condition with theoretically given coefficients and values, *adjustment by correlates*, or, we manage to get an equation for each observation, consequently no equations of condition between several observations This is easily attained by making the number of the elements as large (= m) as may be necessary we may for instance give some values of observations the name of elements This sort of adjustment is called *adjustment by elements* We

shall discuss these two forms in the following chapters XI and XII, first the adjustment by correlates whose rules it is easiest to deduce. In practice we prefer adjustment by correlates when m is nearly as large as n, adjustment by elements when m is small.

XI. ADJUSTMENT BY CORRELATES.

§ 47. We suppose we have ascertained that the whole theory is expressed in the equations $[au] = A, \ldots [cu] = C$, where the adjusted values u of the n observations are the only unknown quantities; we prefer in doubtful cases to have too many equations rather than too few, and occasionally a supernumerary equation to check the computation. The first thing the adjustment by correlates then requires is that the functions $[ao] \ldots [co]$, corresponding to these equations, are made free of one another by the schedule in § 42.

Let $[ao], \ldots [c''o]$ indicate the $n-m$ mutually free functions which we have got by this operation, and let us, beside these, imagine the system of free functions completed by m other arbitrarily selected functions, $[d'''o], \ldots [g^vo]$, representatives of the empiric functions; the adjustment is then principally made by introducing the theoretical values into this system of free functions. It is finally accomplished by transforming back from the free modified functions to the adjusted observations. For this inverse transformation, according to (62), the n equations are

$$o_i = \left\{\frac{a_i}{[aa\lambda_2]}[ao] + \cdots + \frac{c''_i}{[c''c''\lambda_2]}[c''o] + \frac{d'''_i}{[d'''d'''\lambda_2]}[d'''o] + \cdots + \frac{g^v_i}{[g^vg^v\lambda_2]}[g^vo]\right\}\lambda_2(o_i) \quad (73)$$

and according to (65) (compare also (63))

$$\left.\begin{aligned} \lambda_2(o_i) &= \left\{\frac{a_i^2\lambda_2^2(o_i)}{[aa\lambda_2]^2}\lambda_2[ao] + \cdots + \frac{g_i^{v2}\lambda_2^2(o_i)}{[g^vg^v\lambda_2]^2}\lambda_2[g^vo]\right\} \\ &= \left\{\frac{a_i^2}{[aa\lambda_2]} + \cdots + \frac{c_i''^2}{[c''c''\lambda_2]} + \frac{d_i'''^2}{[d'''d'''\lambda_2]} + \cdots + \frac{g_i^{v2}}{[g^vg^v\lambda_2]}\right\}\lambda_2^2(o_i) \end{aligned}\right\} \quad (74)$$

As the adjustment influences only the $n-m$ first terms of each of these equations, we have, because $[au] = A, \ldots [c''u] = C''$, and $\lambda_2[au] = \cdots = \lambda_2[c''u] = 0$,

$$u_i = \left\{\frac{a_i}{[aa\lambda_2]}A + \cdots + \frac{c''_i}{[c''c''\lambda_2]}C'' + \frac{d'''_i}{[d'''d'''\lambda_2]}[d'''o] + \cdots + \frac{g^v_i}{[g^vg^v\lambda_2]}[g^vo]\right\}\lambda_2(o_i) \quad (75)$$

and

$$\lambda_2(u_i) = \left\{\frac{d_i'''^2}{[d'''d'''\lambda_2]} + \cdots + \frac{g_i^{v2}}{[g^vg^v\lambda_2]}\right\}\lambda_2^2(o_i) \quad (76)$$

Consequently

$$o_i - u_i = \lambda_2(o_i)\left\{a_i\frac{[ao]-A}{[aa\lambda_2]} + \dots + c_i''\frac{[c''o]-C''}{[c''c''\lambda_2]}\right\} \tag{77}$$

and

$$\lambda_2(o_i) - \lambda_2(u_i) = \lambda_2^2(o_i)\left\{\frac{a_i^2}{[aa\lambda_2]} + \dots + \frac{c_i''^2}{[c''c''\lambda_2]}\right\} = \lambda_2(o_i - u_i) \tag{78}$$

Thus for the computation of all the differences between the observed and adjusted values of the several observations and the squares of their mean errors, and thereby indirectly for the whole adjustment, we need but use the values and the mean errors of the several observations, the coefficients in the theoretically given functions, and the two values of each of these, namely, the theoretical value, and the value which the observations would give them

The factors in the expression for $o_i - u_i$,

$$K_a = \frac{[ao]-A}{[aa\lambda_2]}, \dots, K_{c''} = \frac{[c''o]-C''}{[c''c''\lambda_2]},$$

which are common to all the observations, are called *correlates*, and have given the method its name The adjusted, improved values of the observations are computed in the easiest way by the formula

$$u_i = o_i - \lambda_2(o_i)\{a_i K_a + \dots + c_i'' K_{c''}\}. \tag{79}$$

By writing the equation (78)

$$\frac{\lambda_2(o_i - u_i)}{\lambda_2(o_i)} = \left\{\frac{a_i^2}{[aa\lambda_2]} + \dots + \frac{c_i''^2}{[c''c''\lambda_2]}\right\}\lambda_2(o_i) \tag{80}$$

and summing up for all values of i from 1 to n, we demonstrate the proposition concerning the sum of the scales discussed in the preceding chapter, viz

$$\left[1 - \frac{\lambda_2(u)}{\lambda_2(o)}\right] = \frac{[aa\lambda_2]}{[aa\lambda_2]} + \dots + \frac{[c''c''\lambda_2]}{[c''c''\lambda_2]} = n - m \tag{81}$$

§ 48. It deserves to be noticed that all these equations are homogeneous with respect to the symbol λ_2. Therefore it makes no change at all in the results of the adjustment or the computation of the scales, if our assumed knowledge of the mean errors in the several observations has failed by a wrong estimate of the unity of the mean errors if only the proportionality is preserved; we can adjust correctly if we know only the relative weights of the observations. The homogeneousness is not broken till we reach the equations of the criticism;

$$\left.\begin{aligned}
&\frac{([ao]-A)^2}{[aa\lambda_1]}+\cdots+\frac{([c''o]-C'')^2}{[c''c''\lambda_1]}=\\
&=K_a^2[aa\lambda_2]+\cdots+K_{c''}^2[c''c''\lambda_2]=\\
&=\left[\frac{(o-u)^2}{\lambda_1(o)}\right]=[(aK_a+\cdots+c''K_{c''})^2\lambda_2(o)]=n-m\pm\sqrt{2(n-m)}
\end{aligned}\right\} \quad (82)$$

It follows that criticism in this form, the "summary criticism", can only be used to try the correctness of the hypothetical unity of the mean errors, or to determine this if it has originally been quite unknown. The special criticism, on the other hand, can, where the series of observations is divided into groups, give fuller information through the sums of squares

$$\Sigma\frac{(o-u)^2}{\lambda_1(o)}=\Sigma\left(1-\frac{\lambda_1(u)}{\lambda_1(o)}\right), \quad (83)$$

taken for each group. We may, for instance, test or determine the unities of the mean errors for one group by means of observations of angles, for another by measurements of distances, etc.

The criticism has also other means at its disposal. Thus the differences $(o-u)$ ought to be small, particularly those whose mean errors have been small, and they ought to change their signs in such a way that approximately

$$\Sigma\frac{o_i-u_i}{\lambda_1(o_i)}=0 \quad (84)$$

for natural or accidentally selected groups, especially for such series of observations as are nearly repetitions, the essential circumstances having varied very little.

If, ultimately, the observations can be arranged systematically, either according to essential circumstances or to such as are considered inessential, we must expect frequent and irregular changes of the signs of $o-u$. If not, we are to suspect the observations of systematical errors, the theory proving to be insufficient.

§ 40. It will not be superfluous to present in the form of a schedule of the adjustment by correlates what has been said here, also as to the working out of the free functions. We suppose then that, among 4 unbound observations o_1, o_2, o_3, and o_4, with the squares on their mean errors $\lambda_1(o_1)$, $\lambda_1(o_2)$, $\lambda_1(o_3)$, and $\lambda_1(o_4)$, there exist relations which can be expressed by the three theoretical equations

$$\begin{aligned}
[au]&=a_1u_1+a_2u_2+a_3u_3+a_4u_4=A\\
[bu]&=b_1u_1+b_2u_2+b_3u_3+b_4u_4=B\\
[cu]&=c_1u_1+c_2u_2+c_3u_3+c_4u_4=C.
\end{aligned}$$

The schedule is then as follows:

		The given			Free functions		
		A	B	C	B'	C'	C''
o_1	$\lambda_2(o_1)$	a_1	b_1	c_1	b'_1	c'_1	c''_1
o_2	$\lambda_2(o_2)$	a_2	b_2	c_2	b'_2	c'_2	c''_2
o_3	$\lambda_2(o_3)$	a_3	b_3	c_3	b'_3	c'_3	c''_3
o_4	$\lambda_2(o_4)$	a_4	b_4	c_4	b'_4	c'_4	c''_4
		$[ao]$	$[bo]$	$[co]$	$[b'o]$	$[c'o]$	$[c''o]$
		$[aa\lambda]$	$[ab\lambda]$	$[ac\lambda]$			
		$[ba\lambda]$	$[bb\lambda]$	$[bc\lambda]$	$[b'b'\lambda]$	$[b'c'\lambda]$	
		$[ca\lambda]$	$[cb\lambda]$	$[cc\lambda]$	$[c'b'\lambda]$	$[c'c'\lambda]$	$[c''c''\lambda]$

$$\beta = \frac{[ba\lambda]}{[aa\lambda]},\quad \gamma = \frac{[ca\lambda]}{[aa\lambda]} \qquad \gamma' = \frac{[c'b'\lambda]}{[b'b'\lambda]}$$

Correlates $\quad K_a = \dfrac{[ao]-A}{[aa\lambda]} \qquad K_{b'} = \dfrac{[b'o]-B'}{[b'b'\lambda]} \qquad K_{c''} = \dfrac{[c''o]-C''}{[c''c''\lambda]}$

Adjusted values				Scales	
o_1-u_1	u_1	$\lambda_2(o_1-u_1)$	$\lambda_2(u_1)$	$1-\lambda_2(u_1)$	$\lambda_2(o_1)$
o_2-u_2	u_2	$\lambda_2(o_2-u_2)$	$\lambda_2(u_2)$	$1-\lambda_2(u_2)$	$\lambda_2(o_2)$
o_3-u_3	u_3	$\lambda_2(o_3-u_3)$	$\lambda_2(u_3)$	$1-\lambda_2(u_3)$	$\lambda_2(o_3)$
o_4-u_4	u_4	$\lambda_2(o_4-u_4)$	$\lambda_2(u_4)$	$1-\lambda_2(u_4)$	$\lambda_2(o_4)$
				$= 3$ as proof	

Criticism	
$(o_1-u_1)^2$	$\lambda_2(o_1)$
$(o_2-u_2)^2$	$\lambda_2(o_2)$
$(o_3-u_3)^2$	$\lambda_2(o_3)$
$(o_4-u_4)^2$	$\lambda_2(o_4)$
Sum for proof and summary criticism	

The free functions are computed by means of

$$\begin{array}{l|l|l}
B' = B-\beta A & C' = C-\gamma A & C'' = C'-\gamma' B' \\
b'_i = b_i-\beta a_i & c'_i = c_i-\gamma a_i & c''_i = c'_i-\gamma' b'_i \\
[b'o] = [bo]-\beta[ao] & [c'o] = [co]-\gamma[ao] & [c''o] = [c'o]-\gamma'[b'o] \\
\cdots & \cdots & \cdots \\
[c'b'\lambda] = [cb\lambda]-\beta[ca\lambda] & [c'c'\lambda] = [cc\lambda]-\gamma[ca\lambda] & [c''c''\lambda] = [c'c'\lambda]-\gamma'[c'b'\lambda]
\end{array}$$

By the adjustment properly so called we compute

$$o_i - u_i = (a_i K_a + b'_i K_{b'} + c''_i K_{c''})\,\lambda_2(o_i)$$

$$\lambda_2(o_i-u_i) = \left(\frac{a_i^2}{[aa\lambda]} + \frac{b'^2_i}{[b'b'\lambda]} + \frac{c''^2_i}{[c''c''\lambda]}\right)(\lambda_2(o_i))^2$$

$$\lambda_2(u_i) = \lambda_2(o_i) - \lambda_2(o_i-u_i),$$

and for the summary criticism

$$K_a^2[aa\lambda_2] + K_{b'}^2[b'b'\lambda_2] + K_{c''}^2[c''c''\lambda_2] = \left[\frac{(o-u)^2}{\lambda_2(o)}\right] = 3 \pm \sqrt{6}$$

In order to get a check we ought further to compute $[au] = A$, $[bu] = B$, and $[cu] = C$, with the values we have found for u_1, u_2, u_3, and u_4. Moreover it is useful to add a superfluous theoretical equation, for instance $[(a+b+c)u] = A+B+C$, through the

computation of the free functions, which is correct only if such a superfluity leads to identical results

§ 50 It is a deficiency in the adjustment by correlates that it cannot well be employed as an intermediate link in a computation that goes beyond it The method is good as far as the determination of the adjusted values of the several observations and the criticism on the same, but no farther We are often in want of the adjusted values with determinations of the mean errors of certain functions of the observations; in order to solve such problems the adjustment by correlates must be made in a modified form The simplest course is, I think, immediately after drawing up the theoretical equations of condition to annex the whole series of the functions that are to be examined, for instance $[do]$, . $[eo]$, and include them in the computation of the free functions In doing so we must take care not to mix up the theoretically and the empirically determined functions, so that the order of the operation must unconditionally give the precedence to the theoretical functions; the others are not made free till the treatment of these is quite finished The functions $[d'''o]$, $[e^{v}o]$, which are separated from these — it is scarcely necessary to mention it — remain unchanged by the adjustment both in value and in mean error And at last the adjusted functions $[du]$, $[eu]$, by retrograde transformation, are determined as linear functions of A, B', C'', $[d'''o]$, . $[e^{v}o]$

Example 1 In a plane triangle each angle has been measured several times, all measurements being made according to the same method, bondfree and with the same (unknown) mean error

for angle A has been found 70° 0′ 5″ as the mean number of 6 measurements
" " B " " " 50° 0′ 3″ " " " " " 10 "
" " C " " " 60° 0′ 2″ " " " " " 15 "

The adjusted values for the angles are then 70°, 50°, and 60°, the mean error for single measurement $= \sqrt{300} = 17''3$, the scales 0 5, 0 3, and 0·2.

Example 2. (Comp example § 42) Five equidistant tabular values, 12, 19, 29, 41, 55, have been obtained by taking approximate round values from an exact table, from which reason their mean errors are all $= \sqrt{\frac{1}{12}}$. The adjustment is performed under the successive hypotheses that the table belongs to a function of the 3rd, 2nd and 1st degree, and the hypothesis of the second degree is varied by the special hypothesis that the 2nd difference is exactly $= 2$, in the following schedule marked (or) The same schedule may be used for all four modifications of the problem, so that in the sums to the right in the schedule, the first term corresponds to the first modification only, and the sum of the two first terms to the second modification.

o	$\lambda_2(o)$	Δ^4	$\nabla\Delta^3$	Δ^2	$(\nabla\Delta^3)'$	$(\Delta^2)' = (\Delta^2)''$	$o-u$	$\lambda_2(o-u)$
		0	0	0 (or 2)	0	0 (or 2)		
12	$\frac{1}{12}$	1	0	0	$-\frac{1}{2}$	$\frac{2}{7}$	$\frac{1}{70}(\ \ 1+\ 7+100$ (or $+20$)),	$\frac{1}{840}(\ 1+\ 7+20)$
19	$\frac{1}{12}$	-4	-1	1	1	$-\frac{1}{7}$	$\frac{1}{70}(-4-14-\ \ 80$ (or -10)),	$\frac{1}{840}(16+28+\ 5)$
29	$\frac{1}{12}$	6	3	-2	0	$-\frac{2}{7}$	$\frac{1}{70}(\ \ 6+\ 0-160$ (or -20)),	$\frac{1}{840}(36+\ 0+20)$
41	$\frac{1}{12}$	-4	-3	1	-1	$-\frac{1}{7}$	$\frac{1}{70}(-4+14-\ \ 80$ (or -10)),	$\frac{1}{840}(16+28+\ 5)$
55	$\frac{1}{12}$	1	1	0	$\frac{1}{2}$	$\frac{2}{7}$	$\frac{1}{70}(\ \ 1-\ 7+100$ (or $+20$)),	$\frac{1}{840}(\ 1+\ 7+20)$
		1	0	2	$-\frac{1}{2}$	$\frac{16}{7}$		
		$\frac{70}{12}$	$\frac{35}{12}$	$-\frac{20}{12}$				
		$\frac{35}{12}$	$\frac{20}{12}$	$-\frac{10}{12}$	$\frac{5}{24}$	0		
		$-\frac{20}{12}$	$-\frac{10}{12}$	$\frac{6}{12}$	0	$\frac{1}{42}$		

$$\beta = \tfrac{1}{2},\quad \gamma = -\tfrac{2}{7} \qquad\qquad \gamma' = 0$$

$$\tfrac{1}{12}K_4 = \tfrac{1}{70},\quad \tfrac{1}{12}K_{\nabla} = -\tfrac{1}{5},\quad \tfrac{1}{12}K_2 = 8 \text{ (or 1)}$$

For the summary criticism

$$\left[\frac{(o-u)^2}{\lambda_2(o)}\right] = \frac{6}{35} + \frac{42}{35} + \frac{7680 \text{ (or } 120)}{35}$$

The hypothesis of the third degree, $\Delta^4 = 0$, where the values of $70u_i$ and their differences are

839 1334 2024 2874 3849

495 690 850 975

195 160 125

-35 -35,

agrees too well with the observations, and must be suspected of being underadjusted, for the sum of the squares of the summary criticism is only

$\frac{6}{35}$, where we might expect $1 \pm \sqrt{2}$.

The hypothesis of the second degree, $\Delta^4 = 0$, $\nabla\Delta^3 = 0$, gives for $70u_i$ and differences

832 1348 2024 2860 3856

516 676 836 996

160 160 160

The adjustment is here good, the sum of the squares is

$\frac{48}{35}$, and we might expect $2 \pm \sqrt{4}$

The hypothesis of the first degree, $\Delta^4 = 0$, $\nabla\Delta^3 = 0$, $\Delta^2 = 0$, gives for the adjusted values and their differences

96 204 312 420 528

108 108 108 108.

The deviations are evidently too large ($o - u$ is $+2.4$, -1.4, -2.2, -1.0, $+2.2$) to be due to the use of round numbers, the sum of the squares is also

$$22.8 \text{ instead of } 3 \pm \sqrt{6},$$

consequently, no doubt, an over-adjustment

The special adjustment of the second degree, $\Delta^4 = 0$, $\nabla\Delta^3 = 0$, and $\Delta^2 = 2$, gives for u_i and its differences,

$$\begin{array}{ccccccccc} 11.6 & & 19.4 & & 29.2 & & 41.0 & & 54.8 \\ & 7.8 & & 9.8 & & 11.8 & & 13.8 & \end{array}$$

The deviations $o - u =$ $+0.4$, -0.4, -0.2, 0.0, $+0.2$

nowhere reach $\frac{1}{2}$, and may consequently be due to the use of round numbers, the sum of the squares

$$4.8 \text{ instead of } 3 \pm \sqrt{6}$$

also agrees very well. Indeed, a constant subtraction of 0.04 from u_i would lead to $(3.4)^2$, $(4.4)^2$, $(5.4)^2$, $(6.4)^2$, and $(7.4)^2$, from which the example is taken

Example 3 Between 4 points on a straight line the 6 distances

$$\begin{array}{ccc} o_{12}, & o_{13}, & o_{14} \\ & o_{23}, & o_{24} \\ & & o_{34} \end{array}$$

are measured with equal exactness without bonds By adjustment we find for instance

$$u_{12} = \tfrac{1}{2} o_{12} + \tfrac{1}{4}(o_{13} - o_{23}) + \tfrac{1}{4}(o_{14} - o_{24}),$$

we notice that every scale $= \frac{1}{2}$ It is recommended actually to work the example by a millimeter scale, which is displaced after the measurement of each distance in order to avoid bonds

XII. ADJUSTMENT BY ELEMENTS.

§ 51 Though every problem in adjustment may be solved in both ways, by correlates as well as by elements, the difficulty in so doing is often very different The most frequent cases, where the number of equations of condition is large, are best suited for adjustment by elements, and this is therefore employed far oftener than adjustment by correlates

The adjustment by elements requires the theory in such a form that *each observation is represented by one equation* which expresses the mean value $\lambda_1(o)$ explicitely as linear functions of unknown values, the "*elements*" x, y, ... z,

$$\left.\begin{array}{l}\lambda_1(o_1) = p_1 x + q_1 y + \quad + r_1 z = u_1 \\ \\ \lambda_1(o_n) = p_n x + q_n y + \quad + r_n z = u_n\end{array}\right\} \quad (85)$$

where the $p, q,$ r are theoretically given All observations are supposed to be unbound.

The problem is then first to determine the adjusted values of these elements $x, y,$ $z,$ after which each of these equations (85), which we call "*equations for the observations*", gives the adjusted value u of the observation

Constantly assuming that $\lambda_2(o)$ is known for each observation, we can from the system (85) deduce the following *normal equations*

$$\left.\begin{array}{l}\left[\frac{p\lambda_1(o)}{\lambda_2(o)}\right] = \left[\frac{pp}{\lambda_2(o)}\right]x + \left[\frac{pq}{\lambda_2(o)}\right]y + \quad + \left[\frac{pr}{\lambda_2(o)}\right]z = \left[\frac{po}{\lambda_2(o)}\right] \\ \left[\frac{q\lambda_1(o)}{\lambda_2(o)}\right] = \left[\frac{qp}{\lambda_2(o)}\right]x + \left[\frac{qq}{\lambda_2(o)}\right]y + \quad + \left[\frac{qr}{\lambda_2(o)}\right]z = \left[\frac{qo}{\lambda_2(o)}\right] \\ \\ \left[\frac{r\lambda_1(o)}{\lambda_2(o)}\right] = \left[\frac{rp}{\lambda_2(o)}\right]x + \left[\frac{rq}{\lambda_2(o)}\right]y + \quad + \left[\frac{rr}{\lambda_2(o)}\right]z = \left[\frac{ro}{\lambda_2(o)}\right]\end{array}\right\} \quad (86)$$

the rule of formation being apparent from the left hand terms Of these normal equations we can prove, first that they, m in number, are suited for the determination of the m elements, so far as these, on the whole, can be determined by the equations (85), and then that the functions of the observations, which form their left hand terms are free of all the theoretical conditions of the problem, so that, as indicated by the last sign of equality in the normal equations, they can and must be determined by the directly observed values o_1 o_n

For if we assume, as to the first proposition, that any of the normal equations can be deduced from the others, so that all the elements cannot be determined by these equations, then there must be m coefficients $h, k,$ $l,$ so that

$$\begin{array}{l}h\left[\frac{pp}{\lambda}\right] + k\left[\frac{qp}{\lambda}\right] + \quad + l\left[\frac{rp}{\lambda}\right] = 0 \\ h\left[\frac{pq}{\lambda}\right] + k\left[\frac{qq}{\lambda}\right] + \quad + l\left[\frac{rq}{\lambda}\right] = 0 \\ \\ h\left[\frac{pr}{\lambda}\right] + k\left[\frac{qr}{\lambda}\right] + \quad + l\left[\frac{rr}{\lambda}\right] = 0\end{array}$$

(λ everywhere used for $\lambda_1(o)$), but if we multiply these again respectively by h, k, l and add, we get

$$\left[\frac{(hp+kq+\quad+lu)^2}{\lambda_1(o)}\right] = 0,$$

that is

$$hp_i + kq_i + \quad + lu_i = 0,$$

so that not only the normal equations, but the very equations for the observations can, consequently, all be written with $m-1$ or a smaller number of elements

But further, the system of functions represented by the normal equations is free of every one of the conditions of the theory The latter we can get by eliminating the elements x, y, z from the equations of the observations (85) But elimination of an element, say for instance x, leads to the functions $p_i\lambda_1(o_k) - p_k\lambda_1(o_i)$, and among the linear functions of these must be found the functions from which not only x but all the other elements are eliminated, and consequently the conditional equations of the theory But it is easily seen that the functions

$$p_i\lambda_1(o_k) - p_k\lambda_1(o_i) \text{ and } \left[\frac{p\lambda_1(o)}{\lambda_1(o)}\right]$$

are mutually free The latter is the left hand side of the normal equation which is particularly aimed at the element x, it is formed by multiplying the equations (85) by the coefficient of x in each, and has the sum of the squares $\left[\frac{pp}{\lambda}\right]$ as the coefficient of this element; it has thus been proved to be free of all the conditions of the theory, and must therefore in the adjustment be computed by the directly observed values, for which reason we have been able in the equations (86) to rewrite the function as $\left[\frac{po}{\lambda_1(o)}\right]$. In the same way we prove that all the other normal equations are free of the theory, each through the elimination from (85) of its particularly prominent element While, in the adjustment by correlates, we exclusively made use of the equations and functions of the theory, we put all these aside in the adjustment by elements, in order to work only with the empirically determined functions which the normal equations represent

The coefficients of the elements in the normal equations are, as it will be seen, arranged in a remarkably symmetrical manner, and each of them has a significance for the problem which it is easy to state

The coefficients in the diagonal line, which are respectively multiplied by the element to which the equation particularly refers, are as sums of squares all positive, and each of them is the square of the mean error for that function of the observations in whose equation it occurs, We have for instance

$$\left[\frac{pp}{\lambda}\right] = \left[\frac{p}{\lambda}\frac{p}{\lambda}\lambda_1(o)\right] = \lambda_1\left[\frac{po}{\lambda}\right]$$

The coefficients outside the diagonal line are identical in pairs, the coefficient of x, $\left[\frac{qp}{\lambda}\right]$ in y's particular equation, is the same as the coefficient of y, $\left[\frac{pq}{\lambda}\right]$ in x's particular equation They show immediately if some of the functions $\left[\frac{po}{\lambda}\right]$, $\left[\frac{qo}{\lambda}\right]$, . . $\left[\frac{ro}{\lambda}\right]$ should happen to be mutually free, If for instance x's function $\left[\frac{po}{\lambda}\right]$ is to be free of y's function $\left[\frac{qo}{\lambda}\right]$, we must have $\left[\frac{p}{\lambda}\,\frac{q}{\lambda}\cdot\lambda_2(o)\right] = \left[\frac{pq}{\lambda}\right] = 0$

§ 52. If now the elements have been selected in such a convenient way that all these sums of the products vanish, and the normal equations consequently appear in the special form

$$\left.\begin{array}{rl}\left[\frac{pp}{\lambda}\right]x \qquad\qquad & = \left[\frac{po}{\lambda}\right] \\ \left[\frac{qq}{\lambda}\right]y \qquad & = \left[\frac{qo}{\lambda}\right] \\ \cdot & \\ \left[\frac{rr}{\lambda}\right]z & = \left[\frac{ro}{\lambda}\right]\end{array}\right\} \tag{87}$$

then they offer us directly the solution of the problem of adjustment The adjusted values for the elements are

$$x = \left[\frac{po}{\lambda}\right] : \left[\frac{pp}{\lambda}\right],\quad y = \left[\frac{qo}{\lambda}\right] : \left[\frac{qq}{\lambda}\right],\quad \ldots\quad z = \left[\frac{ro}{\lambda}\right] : \left[\frac{rr}{\lambda}\right], \tag{88}$$

and the squares of the mean errors

$$\lambda_2(x) = \left[\frac{pp}{\lambda}\right]^{-1},\quad \lambda_2(y) = \left[\frac{qq}{\lambda}\right]^{-1},\quad \ldots\quad \lambda_2(z) = \left[\frac{rr}{\lambda}\right]^{-1}, \tag{89}$$

and from these we can then compute both the adjusted value and its λ_2 for every linear function of the elements, because these are mutually free functions In particular from the equations (85),

$$u_i = p_i x + q_i y + \ldots + r_i z,$$

we can compute the adjusted values u_i of the observations, then from (86) the squares of the mean errors $\lambda_2(u_i)$, and also the law of errors for every function of observations and elements.

§ 53. In ordinary cases a transformation of the system of elements is required It is required for the solution of the normal equations in order to find the values of the elements, but we must remember that we have here a double problem, as it is also our object to free the transformed elements so that they may be used for determinations of the mean errors. The transformation therefore cannot be selected so arbitrarily as in analogous problems of pure mathematics; yet there is a multiplicity of possibilities, and

in many special cases radical changes can lead to very beautiful solutions (see § 02) The first thing, however is to secure a method which may be always applied, and this must be selected in such a way that the elements are eliminated one by one, so that the later computation of them is prepared, and moreover, constantly, in such a way that freedom is attained

This can, if we commence for instance by eliminating the element x, be attained in the following way The normal equation which particularly refers to x,

$$\left[\frac{pp}{\lambda}\right]x + \left[\frac{pq}{\lambda}\right]y + \quad + \left[\frac{pr}{\lambda}\right]z = \left[\frac{po}{\lambda}\right], \tag{90}$$

and which will be put aside to be used later on for the computation of x, is multiplied by such factors, viz $\varphi = \left[\frac{qp}{\lambda}\right] : \left[\frac{pp}{\lambda}\right]$, $\quad \omega = \left[\frac{rp}{\lambda}\right] : \left[\frac{pp}{\lambda}\right]$, that x vanishes when the products are respectively subtracted from the other normal equations, but it must be remembered that we are not allowed to multiply the latter by any factor The equation for x can then be written

$$\xi = x + \varphi y + , \quad + \omega z = \left[\frac{po}{\lambda}\right] : \left[\frac{pp}{\lambda}\right] \tag{91}$$

where $\lambda_1(\xi) = \left[\frac{pp}{\lambda}\right]^{-1}$

The functions in the other equations

$$\left[\frac{qo}{\lambda}\right] - \varphi\left[\frac{po}{\lambda}\right] = \left[\frac{(q-\varphi p)o}{\lambda}\right], \qquad \left[\frac{ro}{\lambda}\right] - \omega\left[\frac{po}{\lambda}\right] = \left[\frac{(r-\omega p)o}{\lambda}\right]$$

become, by this means, not only independent of x but also free of $\left[\frac{po}{\lambda}\right]$ or of ξ, for

$$\left[\frac{p}{\lambda}\;\frac{q-\varphi p}{\lambda}\;\lambda_1(o)\right] = \left[\frac{pq}{\lambda}\right] - \varphi\left[\frac{pp}{\lambda}\right] = 0, \text{ etc}$$

The equations which in a double sense have been freed from p, get exactly the same characteristic functional form as the normal equations had If we write

$$q'_i = q_i - \varphi p_i, \qquad r'_i = r_i - \omega p_i, \tag{92}$$

so that the equations for the observations become

$$p_i\xi + q'_i y + \quad + r'_i z = u_i,$$

we not only get, as we see at once,

$$\left[\frac{q'o}{\lambda}\right] = \left[\frac{qo}{\lambda}\right] - \varphi\left[\frac{po}{\lambda}\right], \qquad \left[\frac{r'o}{\lambda}\right] = \left[\frac{ro}{\lambda}\right] - \omega\left[\frac{po}{\lambda}\right], \tag{93}$$

but also

$$\left.\begin{aligned}\left[\frac{q'q'}{\lambda}\right] &= \left[\frac{qq}{\lambda}\right] - \varphi\left[\frac{qp}{\lambda}\right], \qquad & \left[\frac{q'r'}{\lambda}\right] &= \left[\frac{qr}{\lambda}\right] - \omega\left[\frac{qp}{\lambda}\right] \\ \left[\frac{r'q'}{\lambda}\right] &= \left[\frac{rq}{\lambda}\right] - \varphi\left[\frac{rp}{\lambda}\right], \qquad & \left[\frac{r'r'}{\lambda}\right] &= \left[\frac{rr}{\lambda}\right] - \omega\left[\frac{rp}{\lambda}\right]\end{aligned}\right\} \tag{94}$$

Hence we proceed exactly in the same way from this first stage of the transformation of the normal equations

$$\left.\begin{aligned}\left[\frac{q'q'}{\lambda}\right]y + \quad + \left[\frac{q'r'}{\lambda}\right]z &= \left[\frac{q'o}{\lambda}\right] \\ \left[\frac{r'q'}{\lambda}\right]y + \quad + \left[\frac{r'r'}{\lambda}\right]z &= \left[\frac{r'o}{\lambda}\right]\end{aligned}\right\} \tag{95}$$

using, for instance, the first of them for the elimination of the element y If

$$\omega' = \left[\frac{r'q'}{\lambda}\right] \ \left[\frac{q'q'}{\lambda}\right],$$

y is replaced by

$$\eta = y + \quad + \omega' z = \left[\frac{q'o}{\lambda}\right] \ \left[\frac{q'q'}{\lambda}\right], \tag{96}$$

which is free of the element ξ, and for which we have

$$\lambda_1(\eta) = \left[\frac{q'q'}{\lambda}\right]^{-1} \tag{97}$$

By means of ω' and corresponding coefficients we have, analogously to (93) and (94),

$$\left[\frac{r''o}{\lambda}\right] = \left[\frac{r'o}{\lambda}\right] - \omega'\left[\frac{q'o}{\lambda}\right], \qquad \left[\frac{r''r''}{\lambda}\right] = \left[\frac{r'r'}{\lambda}\right] - \omega'\left[\frac{r'q'}{\lambda}\right],$$

which are independent of any special computation of the coefficients r''

Continuing in this way, till we have obtained a set consisting only of free functions, we find, consequently, just a system of elements, ξ, η, ζ, which possess the abovementioned desired property, its normal equations being of the same form as (87), viz

$$\left.\begin{aligned}\left[\frac{pp}{\lambda}\right]\xi \qquad\qquad &= \left[\frac{po}{\lambda}\right] \\ \left[\frac{q'q'}{\lambda}\right]\eta \qquad &= \left[\frac{q'o}{\lambda}\right] \\ \left[\frac{r''r''}{\lambda}\right]\zeta &= \left[\frac{r''o}{\lambda}\right]\end{aligned}\right\} \tag{98}$$

With these elements the equations for the adjusted values of the several observations become

$$p_i\xi + q_i'\eta + \quad + r_i^{\nu}\zeta = u_i, \tag{99}$$

and for the squares of their mean errors

$$p_i^2\left[\frac{pp}{\lambda}\right]^{-1} + q_i'^2\left[\frac{q'q'}{\lambda}\right]^{-1} + \quad + r_i^{\nu 2}\left[\frac{r^{\nu}r^{\nu}}{\lambda}\right]^{-1} = \lambda_2(u_i). \tag{100}$$

If we want to compute adjusted values and mean errors for the original elements or functions of the same, the means of so doing is given by the equations of transformation

$$\left.\begin{array}{r} x + \varphi y + \quad + \omega z = \xi \\ y + \quad + \omega' z = \eta \\ \\ z = \zeta \end{array}\right\} \tag{101}$$

or by (90), the first equation (95) and the last of (98), being identical with (101). For not only the original elements x, y, z are easily computed by these, but also the coefficients in the inverse transformation

$$\left.\begin{array}{l} x = \xi + \alpha\eta + \quad + \gamma\zeta \\ y = \quad\quad \eta + \quad + \delta\zeta \\ z = \quad\quad\quad\quad\quad \zeta \end{array}\right\} \tag{102}$$

Now, if F is a given linear function of x, y, z, then by obvious numerical operations we get an expression for it,

$$F = a\xi + b\eta + \quad + d\zeta,$$

and for the square of its mean error we get

$$\lambda_2(F) = a^2\left[\frac{pp}{\lambda}\right]^{-1} + b^2\left[\frac{q'q'}{\lambda}\right] + \quad + d^2\left[\frac{r^{\nu}r^{\nu}}{\lambda}\right]$$

If for special criticism we want the computation of $\lambda_2(u_i)$ for many observations, we may take advantage of transforming the equations of observations, computing their coefficients by (102), or

$$q_i' = q_i - \varphi p_i, \qquad r_i' = r_i - \omega p_i$$
$$r_i'' = r_i' - \omega' q_i',$$

but we remember that q', r'' are quite superfluous for the coefficients of (95).

§ 54. In the theory of the adjustment by elements we must not overlook the proposition concerning the computation of the minimum sum of squares for the benefit of the summary criticism as well as for checking our computation. We are able to compute the sum $\left[\frac{(o-u)^2}{\lambda}\right]$, which is to approach the value $n - m$, as soon as we have found only the elements, without being obliged to know the adjusted values for the separate

observations And this computation can be performed, not only for the legitimate adjustment, but for any values whatever of the elements It is easiest to show this for transformed elements, $\xi_1, \eta_1, \dots \zeta_1$ The values for the observations corresponding to these must be computed by (99)

$$p_i\xi_1 + q_i'\eta_1 + \dots + r_i^{\nu}\zeta_1 = v_i$$

From this we get

$$\left[\frac{(o-v)^2}{\lambda_1(o)}\right] = \left[\frac{oo}{\lambda}\right] - 2\left[\frac{po}{\lambda}\right]\xi_1 - 2\left[\frac{q'o}{\lambda}\right]\eta_1 - \dots - 2\left[\frac{r^{\nu}o}{\lambda}\right]\zeta_1 + \left[\frac{pp}{\lambda}\right]\xi_1^2 + \left[\frac{q'q'}{\lambda}\right]\eta_1^2 + \dots + \left[\frac{r^{\nu}r^{\nu}}{\lambda}\right]\zeta_1^2 \quad (103)$$

If we here substitute for $\left[\frac{po}{\lambda}\right], \left[\frac{q'o}{\lambda}\right], \dots \left[\frac{r^{\nu}o}{\lambda}\right]$ their values in terms of the elements $\xi, \eta, \dots \zeta$, of the legitimate adjustment, we find from the equations (98)

$$\left[\frac{(o-v)^2}{\lambda_1(o)}\right] = \left[\frac{oo}{\lambda}\right] + \left[\frac{pp}{\lambda}\right]((\xi_1-\xi)^2-\xi^2) + \left[\frac{q'q'}{\lambda}\right]((\eta_1-\eta)^2-\eta^2) + \dots + \left[\frac{r^{\nu}r^{\nu}}{\lambda}\right]((\zeta_1-\zeta)^2-\zeta^2) \quad (104)$$

It is evident from this that the condition of minimum is $\xi_1 = \xi$, $\eta_1 = \eta$, $\zeta_1 = \zeta$ The minimum sum of squares is therefore obtained only by the determination of the functions that are free of the theory, by means of their directly observed values And for this minimum

$$\left[\frac{(o-u)^2}{\lambda_1(o)}\right] = \left[\frac{oo}{\lambda}\right] - \left[\frac{pp}{\lambda}\right]\xi^2 - \left[\frac{q'q'}{\lambda}\right]\eta^2 - \dots - \left[\frac{r^{\nu}r^{\nu}}{\lambda}\right]\zeta^2 = \quad (105)$$

$$= \left[\frac{oo}{\lambda}\right] - \left[\frac{po}{\lambda}\right]\xi - \left[\frac{q'o}{\lambda}\right]\eta - \dots - \left[\frac{r^{\nu}o}{\lambda}\right]\zeta = \quad (106)$$

$$= \left[\frac{oo}{\lambda}\right] - \frac{\left[\frac{po}{\lambda}\right]^2}{\left[\frac{pp}{\lambda}\right]} - \frac{\left[\frac{q'o}{\lambda}\right]^2}{\left[\frac{q'q'}{\lambda}\right]} - \dots - \frac{\left[\frac{r^{\nu}o}{\lambda}\right]^2}{\left[\frac{r^{\nu}r^{\nu}}{\lambda}\right]} \quad (107)$$

It deserves to be noticed that the middle one of these expressions holds good, in unchanged form, also of the original, not transformed elements and coefficients We have

$$\left[\frac{(o-u)^2}{\lambda_1(o)}\right] = \left[\frac{oo}{\lambda}\right] - \left[\frac{po}{\lambda}\right]x - \left[\frac{qo}{\lambda}\right]y - \dots - \left[\frac{ro}{\lambda}\right]z, \quad (108)$$

which is easily proved by substituting in (106) the values obtained from (101) The equation is particularly valuable as a check on the accuracy of our computation

§ 55 In going through the theory of adjustment by elements here developed, it will be seen that a very essential part of the work, viz the computation of the trans-

formed values of the coefficients in the equations for the several observations, may nearly always be dispensed with. The sums of the squares, $\left[\frac{qq}{\lambda}\right]$, and the sums of the products, $\left[\frac{qr}{\lambda}\right]$, must be transformed; but they are in themselves sufficient for the determination of the transformations, and by their help we find values and mean errors for the elements, first the transformed ones, but indirectly also the original ones. The adjusted values $u_1 \ldots u_n$ of the observations can, consequently, also be computed without any knowledge of $q_i' \ldots r_i' \ldots r_i''$. Only for the computation of $\lambda_2(u_1) \ldots \lambda_2(u_n)$, consequently for a special criticism, we cannot escape the often considerable work which is necessary for the purpose.

For the summary criticism by $\left[\frac{(o-u)^2}{\lambda_1(o)}\right] = n - m \pm \sqrt{2(n-m)}$, we can even, as we have seen, dispense with the after computation of the several observations by means of the elements. We ought, however, to restrict the work of adjustment so far only, when the case is either very difficult or of slight importance, for this minimum sum of squares is generally computed much more sharply, and always with much greater certainty, directly by o_i, u_i, and $\lambda_1(o)$, than by the formulæ (105), (106), and (107).

Add to this, that the special criticism does not exclusively rest on $\lambda_2(u_i)$ and the scales $1 - \frac{\lambda_2(u)}{\lambda_1(o)}$, but that the very deviations $o_i - u_i$, when they are arranged according to the more or less essential circumstances of the observations, are even a main point in the criticism. Systematical errors, especially inaccuracies or defects in hypotheses and theories, will betray themselves in the surest and easiest way by the progression of the errors; regular variation in $o - u$ as a function of some circumstance, or mere absence of frequent changes of signs, will disclose errors which might remain hidden by the check according to $\Sigma \frac{(o-u)^2}{\lambda_1(o)} = \Sigma\left(1 - \frac{\lambda_2(u)}{\lambda_1(o)}\right)$, and such progression in the errors may, we know, even be used to indicate how we ought to try to improve the defective theory.

§ 56. By series of adjustment (compare Dr. J. P. Gram, Udjevningsrækker, Kjøbenhavn 1879, and Crelle's Journal vol. 94), i. e. where the theory gives the observations in the form of a series with an indeterminate (infinite) number of terms, each term being multiplied by an unknown factor, an element, and where consequently adjustment by elements must be employed, the criticism gets the special task of indicating how many (or which) terms of the series we are to include in the adjustment. Formula (107) furnishes us with the means of doing this:

$$\left[\frac{(o-u)^2}{\lambda_1(o)}\right] = \left[\frac{oo}{\lambda}\right] - \Sigma \frac{\left[\frac{r^{\nu}o}{\lambda}\right]^2}{\left[\frac{r^{\nu}r^{\nu}}{\lambda}\right]} = n - m \pm \sqrt{2(n-m)}$$

For the m terms in the series, which is here indicated by Σ, correspond, each of them, to an element, consequently to one of the terms of the series of adjustment. For each term we take into this, the right side of the equation of criticism is diminished by about a unity, the result of the criticism, consequently, becomes more favourable if we leave out all the terms for which $\left[\frac{r^\nu o}{\lambda}\right]^2 \left[\frac{r^\nu r^\nu}{\lambda}\right]^{-1} < 1$. If we retain any terms which essentially fall under this rule, the adjustment becomes an under-adjustment; if, on the other hand, we leave out terms for which $\left[\frac{r^\nu o}{\lambda}\right]^2 \left[\frac{r^\nu r^\nu}{\lambda}\right]^{-1} > 1$, we make ourselves guilty of an over-adjustment.

Example 1. The five-place logarithms in a table are looked upon as mutually unbound observations for which the mean error is constantly $\sqrt{\frac{1}{12}}$ of the fifth decimal place. The "observations", log 795, log 796, log 797, log 798, log 799, log 800, log 801, log 802, log 803, log 804, and log 805, are to be adjusted as an integral function of the second degree

$$\log(800+t) = x' + y't + zt^2.$$

In order to reckon with small integral numbers, we subtract before the adjustment $2.90309 + 0.00054\,t$, both from the observations and from the formulæ. Taking 0.00001 as our unity, we have then the equations for the observations

$$\begin{aligned}
-2 &= x - 5y + 25z \\
-2 &= x - 4y + 16z \\
-1 &= x - 3y + \ \ 9z \\
-1 &= x - 2y + \ \ 4z \\
0 &= x - 1y + \ \ 1z \\
0 &= x \\
0 &= x + 1y + \ \ 1z \\
0 &= x + 2y + \ \ 4z \\
1 &= x + 3y + \ \ 9z \\
1 &= x + 4y + 16z \\
1 &= x + 5y + 25z
\end{aligned}$$

From this we get $\left[\frac{oo}{\lambda}\right] = 156$, and the normal equations

$$\begin{aligned}
-36 &= 132x + \quad 0y + 1320z \\
420 &= \quad 0x + 1320y + \quad 0z \\
-540 &= 1320x + \quad 0y + 23496z
\end{aligned}$$

The element y is consequently immediately free of x and z, but the latter must be made

free of one another, which is done by multiplying the first equation by 10 and subtracting it from the third The transformation into free functions then only requires $\xi = x + 10z$ substituted for x, and we have

$$\begin{aligned} -\ 36 &= \ \ 132\,\xi\,, \\ 420 &= \ \ 1320\,y\,, \\ -180 &= 10290\,z\,, \end{aligned}$$

consequently,

$$\begin{aligned} \xi &= -0\ 2727\,, & \lambda_1(\xi) &= 1 \quad 132 = 390\ ,\ 51480 = 007576 \\ y &= \quad 0\ 3182\,, & \lambda_1(y) &= 1 \quad 1320 = \ \ 39\ \ 51480 = 000758 \\ z &= -0\ 0175\,, & \lambda_1(z) &= 1 \ \ 10296 = \quad 5\ \ 51480 = 000097 \end{aligned}$$

The mean error of y is consequently $\pm 0\ 0275$, and that of z $\pm 0\ 0099$ The element x is found by $x = \xi - 10z = -0\ 0977$, to which corresponds $\lambda_1(x) = \lambda_1(\xi) + 100\lambda_1(z) = 0\ 0173 = (0{\cdot}1315)^2$ For $\log 800$ we find thus $2\ 9030890 \pm 0\ 0000013$, and the corresponding difference of the table is $54\ 318 \pm 0\ 028$

For the sum of the squares of the deviations we have, according to (105)—(107),

$$\left[\frac{(o-u)^2}{\lambda_1(o)}\right] = 156 - 9\ 82 - 133\ 64 - 3\ 15 = 9\ 39\,,$$

which shows that the term of the second degree contributes somewhat to the goodness of the adjustment This sum of squares ought, according to the number of the observations and the elements, to be $11 - 3 = 8$, with a mean uncertainty of ± 4

The best formula for computing the adjusted values of the several observations and their mean errors is $u_t = \xi + yt + z(t^2 - 10)$, which gives

	u	$o-u$	$(o-u)^2$		$\lambda_1(u)$	Scale
log 795	= 2 9003886	+ 12	0144	390 + 39 25 + 5 225 = 2490	0484	419
log 796	= 2 9009136	− 36	1296	390 + 39 16 + 5 36 = 1194	0232	722
log 797	= 2 9014580	+ 20	0400	390 + 39 9 + 5 1 = 746	0145	826
log 798	= 2 9020019	− 19	0361	390 + 39 4 + 5 36 = 726	0141	831
log 799	= 2 9025457	+ 48	1840	390 + 39 · 1 + 5 81 = 834	0162	808
log 800	= 2 9030890	+ 10	0100	390 + 39 0 + 5 100 = 890	0173	792
log 801	= 2 9036321	− 21	0441	390 + 39 1 + 5 81 = 834	0162	808
log 802	= 2 9041747	− 47	2209	390 + 39 4 + 5 36 = 726	0141	·831
log 803	= 2 9047170	+ 30	0900	390 + 39 9 + 5 1 = 746	0145	·826
log 804	= 2 9052590	+ 10	0100	390 + 39 · 16 + 5 36 = 1194	0232	722
log 805	= 2 9058006	− ·06	0036	390 + 39 · 25 + 5 225 = 2490	·0484	419
			7886	12870		8·000

Both the checks agree the sum of squares is $12 \times 0.7836 = 9.40$, and the sum of the scales is $11-3$

It ought to be noticed that the adjustment gives very accurate results throughout the greater part of the interval, with the exception of the beginning and the end The exactness, however, is not greatest in the middle, but near the 1st and the 3rd quarter

Example 2 A finite, periodic function of one single essential circumstance, an angle V, is supposed to be the object of observation The theory, consequently, has the form

$$o_v = c_0 + c_1 \cos V + s_1 \sin V + c_2 \cos 2V + s_2 \sin 2V +$$

We assume that there are n unbound, equally exact observations for a series of values of V, whose difference is constant and $= \frac{2\pi}{n}$, for instance for $V = 0, 60°, 120°, 180°, 240°, 300°$ Show that the normal equations are here originally free, and that they admit of an exceedingly simple computation of each isolated term of the periodic series

Example 3 Determine the abscissæ for 4 points on a straight line whose mutual distances are measured equally exactly, and are unbound (Comp Adjustment by Correlates, Example 3, and § 60)

Example 4 Three unbound observations must according to theory, depend on two elements, so that

$$o_1 = x^2, \quad \lambda_1(o_1) = 1$$
$$o_2 = xy, \quad \lambda_1(o_2) = \tfrac{1}{2}$$
$$o_3 = y^2, \quad \lambda_1(o_3) = 1$$

The theory, therefore, does not give us equations of the linear form This may be produced in several ways, most simply by the common method of presupposing approximate values of both elements, the known a for x and b for y, and considering the corrections ξ and η to be the elements of the adjustment We therefore put $x = a + \xi$, and $y = b + \eta$. Rejecting terms of the 2nd degree, we get the equations of the observations

$$o_1 - a^2 = 2a\xi$$
$$o_2 - ab = b\xi + a\eta$$
$$o_3 - b^2 = 2b\eta,$$

where the middle equation has still double weight The normal equations are

$$2a(o_1 - a^2) + 2b(o_2 - ab) = (4a^2 + 2b^2)\xi + 2ab\eta$$
$$2a(o_2 - ab) + 2b(o_3 - b^2) = \qquad 2ab\xi + (4b^2 + 2a^2)\eta,$$

ξ is consequently not free of η, but we find

$$2ax = o_1 + a^2 - \frac{b^2(b^2 o_1 - 2abo_2 + a^2 o_3)}{(a^2+b^2)^2}, \quad \lambda_2(x) = \frac{a^2+2b^2}{4(a^2+b^2)^2}$$

$$2by = o_3 + b^2 - \frac{a^2(b^2 o_1 - 2abo_2 + a^2 o_3)}{(a^2+b^2)^2}, \quad \lambda_2(y) = \frac{2a^2+b^2}{4(a^2+b^2)^2},$$

For the adjusted value u_2 of the middle observation we have

$$(a^2+b^2)^2 u_2 = ab^3 o_1 + (a^4+b^4) o_2 + a^3 b o_3 , \quad \lambda_2(u_2) = \frac{1}{2} \frac{a^4+b^4}{(a^2+b^2)^2}$$

If we had transformed the elements (comp § 62) by putting

$$\xi = a\zeta - bv$$
$$\eta = b\zeta + av ,$$

or

$$x = a(1+\zeta) - bv$$
$$y = b(1+\zeta) + av ,$$

we should have obtained free normal equations

$$2(a^2 o_1 + 2ab o_2 + b^2 o_3) - 2(a^2+b^2)^2 = 4(a^2+b^2)^2 \zeta$$
$$2(-ab o_1 + (a^2-b^2) o_2 + ab o_3) \qquad = 2(a^2+b^2)^2 v$$

If we had placed absolute confidence in the adjusting principle of the sum of squares as a minimum, a solution might have been founded on

$$(o_1 - a^2)^2 + 2(o_2 - ab)^2 + (o_3 - b^2)^2 = \min$$

The conditions of minimum are

$$\frac{1}{4} \frac{d \min}{da} = (o_1 - a^2) a + (o_2 - ab) b = 0$$
$$\frac{1}{4} \frac{d \min}{db} = (o_2 - ab) a + (o_3 - b^2) b = 0$$

The solution with respect to a and b is not very difficult We see for instance immediately that

$$(o_1 - a^2)(o_3 - b^2) = (o_2 - ab)^2$$

or

$$o_1 o_3 - o_2^2 = b^2 o_1 - 2ab o_2 + a^2 o_3$$

Still better is it to introduce $s^2 = a^2 + b^2$, by which the equations become

$$(o_1 - s^2) a + o_2 b = 0$$
$$o_2 a + (o_3 - s^2) b = 0 ,$$

consequently,

$$s^4 - s^2(o_1 + o_3) + o_1 o_3 - o_2^2 = 0$$
$$\left(s^2 - \frac{o_1 + o_3}{2}\right)^2 = \left(\frac{o_1 - o_3}{2}\right)^2 + o_2^2$$

If the errors in o_1, o_2, and o_3 are not large, $o_1 o_3 - o_2^2$ must be small; one of the two values of s^2 must then be small, the other nearly equal to $o_1 + o_3$. only the latter can be used

Further, we get

$$-\frac{a}{b} = \frac{o_2}{o_1 - s^2} = \frac{o_3 - s^2}{o_2}$$

$$\left(\frac{a}{b}\right)^2 = \frac{o_3 - s^2}{o_1 - s^2}$$

$$a^2 = \frac{(o_3 - s^2)s^2}{o_1 + o_3 - 2s^2}, \quad b^2 = \frac{(o_1 - s^2)s^2}{o_1 + o_3 - 2s^2}$$

In this way we avoid guessing at approximate values (for which otherwise we should perhaps have taken $a^2 = o_1$ and $b^2 = o_3$). The values which we have here found for a^2 and b^2, and to which may be added

$$-ab = \frac{o_2 s^2}{o_1 + o_3 - 2s^2},$$

are really exact; and if we substitute them in the above normal equations, we get $\xi = 0$ and $\eta = 0$.

Even when, as in this case, the theory is not linear, it is not unusual for the sum of the squares to be a minimum. Caution, however, is necessary; particularly, it may happen that the sum of the squares becomes a maximum for the found elements, or for some of them.

We may also in another way make the equations of this example linear, namely, by considering the logarithms of o_1, o_2, o_3 as the observed quantities, and finding the logarithms of the elements from the equations which will then be linear:

$$\begin{aligned}
\log o_1 &= 2\log x \\
\log o_2 &= \log x + \log y \\
\log o_3 &= \qquad\quad 2\log y.
\end{aligned}$$

In this way we throw the difficulty over upon the squares of the mean errors. As

$$\log(z + dz) = \log z + \frac{dz}{z},$$

we may approximately take

$$\lambda_2(\log z) = \frac{1}{z^2}\lambda_2(z).$$

If a and b also here indicate approximate values of x and y, the weights of the 3 equations, respectively, become proportional to a^4, $2a^2b^2$, and b^4. Thus we find the normal equations

$$\begin{aligned}
2a^4\log o_1 + 2a^2b^2\log o_2 &= (4a^4 + 2a^2b^2)\log x + 2a^2b^2\log y \\
2a^2b^2\log o_2 + 2b^4\log o_3 &= \qquad 2a^2b^2\log x + (4b^4 + 2a^2b^2)\log y,
\end{aligned}$$

which give the simple results

$$2\log x = \log o_1 - \left(\frac{b^2}{a^2+b^2}\right)^2 \log\frac{o_1 o_2}{o_3^2}\,,\quad \lambda_2(\log x) = \frac{a^2+2b^2}{4a^2(a^2+b^2)^2}$$

$$2\log y = \log o_2 - \left(\frac{a^2}{a^2+b^2}\right)^2 \log\frac{o_1 o_2}{o_3^2}\,,\quad \lambda_2(\log y) = \frac{2a^2+b^2}{4b^2(a^2+b^2)^2}$$

This solution agrees only approximately with the preceding one It might seem for a moment that, in this way, we might do without the supposition of approximate values for the elements, but this is far from being the case For the sake of the weights we must, with the same care, demand that a and x, as also b and y, agree, and we must repeat the adjustment till the squares of the mean errors get the *theoretically* correct values And then it is only a necessary, but not a sufficient condition, that $x-a$ and $y-b$ are small Unless the exactness of the observations is also so great that the mean errors of o_i are small in proportion to o_i itself, the laws of errors of the logarithms cannot be considered typical at the same time as those of the observations themselves

Example 5 The co-ordinates of four points in a circle are observed with equal mean errors and without bonds $x_1 = 20$, $y_1 = 10$, $x_2 = 16$, $y_2 = 18$; $x_3 = 8$, $y_3 = 17$, and $x_4 = 2$, $y_4 = 4$ In the adjustment for the co ordinates a and b of the centre and the radius r, we cannot use the common form of the equations

$$(x-a)^2 + (y-b)^2 = r^2\,,$$

because it embraces more than *one* observed quantity besides the elements In order to obtain the separation of the observations necessary for adjustment by elements, we must add a supplementary element, or parameter, V_i for each point, writing for instance

$$x_i = a + r\cos V_i\,,\quad y_i = b + r\sin V_i$$

As the equations are not linear we must work by successive corrections Δa, Δb, Δr, ΔV_i of the elements, of which the first approximate system can be obtained by ordinary computation from 3 points For the theoretical corrections Δx_i and Δy_i of the co-ordinates we get by differentiation of the above equations

$$\Delta x_i = \Delta a + \Delta r \cdot \cos V_i - \Delta V_i \cdot r\sin V_i$$
$$\Delta y_i = \Delta b + \Delta r \cdot \sin V_i + \Delta V_i \cdot r\cos V_i$$

These equations for the observations lead us to a system of seven normal equations By the "method of partial elimination" (§ 61) these are not difficult to solve, but here the simplicity of the problem makes it possible for us immediately to discover the artifice We know that every transformation of equally well observed rectangular co-ordinates results in free functions The radial and the tangential corrections

$$\Delta x_i \cos V_i + \Delta y_i \sin V_i = \Delta n_i$$

and

$$\Delta x_i \sin V_i - \Delta y_i \cos V_i = \Delta t_i$$

can, consequently, here be taken directly for the mean values of corrections of observed quantities, and as only the four equations

$$\Delta t_i = \Delta a \sin V_i - \Delta b \cos V_i - r\,\Delta V_i$$

contain the four corrections ΔV_i of the parameters, they can be legitimately reserved for the successive corrections of the elements. In this way

$$\Delta n_i = \Delta a \cos V_i + \Delta b \sin V_i + \Delta r$$

with equal mean errors, $\lambda_1(n) = \lambda_1(x) = \lambda_1(y)$, are the "equations for the observations" of this adjustment, and give the three normal equations

$$\begin{aligned}
[\Delta n \cos V] &= \Delta a [\cos^2 V] + \Delta b [\cos V \sin V] + \Delta r [\cos V] \\
[\Delta n \sin V] &= \Delta a [\cos V \sin V] + \Delta b [\sin^2 V] + \Delta r [\sin V] \\
[\Delta n] &= \Delta a [\cos V] + \Delta b [\sin V] + \Delta r\, 4.
\end{aligned}$$

In the special case under consideration, we easily see that the first, second, and fourth point lie on the circle with $r = 10$, whose centre has the co-ordinates $a = 10$ and $b = 10$; the parameters are consequently

$$V_1 = 0°0'0, \quad V_2 = 53°7'8, \quad V_3 = 135°0'0, \quad \text{and} \quad V_4 = 216°52'2$$

For the third point the computed co-ordinates are $x_3 = 2.9290$ and $y_3 = 17.0710$, consequently, $\Delta x_3 = +0.0710$ and $\Delta y_3 = -0.0710$, $\Delta t_3 = 0$, and $\Delta n_3 = -0.1005$; all other differences $\Delta x_i = 0$ and $\Delta y_i = 0$. The "equations for the observations" are

$$\begin{aligned}
1.0000\,\Delta a + 0.0000\,\Delta b + 1.0000\,\Delta r &= 0.0000 \\
0.6000\,\Delta a + 0.8000\,\Delta b + 1.0000\,\Delta r &= 0.0000 \\
-0.7071\,\Delta a + 0.7071\,\Delta b + 1.0000\,\Delta r &= -0.1005 \\
-0.8000\,\Delta a - 0.6000\,\Delta b + 1.0000\,\Delta r &= 0.0000
\end{aligned}$$

The normal equations are

$$\begin{array}{rcl|l}
 & & & \text{For checking} \\
2.5000\,\Delta a + 0.4600\,\Delta b + 0.0929\,\Delta r &=& +0.0710 & +0.0002 \\
0.4600\,\Delta a + 1.5000\,\Delta b + 0.9071\,\Delta r &=& -0.0710 & 0.0000 \\
R = 0.0929\,\Delta a + 0.9071\,\Delta b + 4.0000\,\Delta r &=& -0.1005 & 0.0000 \\
\text{By elimination of } \Delta r \text{ we get} & & & \\
2.4978\,\Delta a + 0.4890\,\Delta b &=& +0.0733 & +0.0001 \\
B = 0.4890\,\Delta a + 1.2943\,\Delta b &=& -0.0482, & -0.0001 \\
\text{and by eliminating } \Delta b & & & \\
A = +2.3490\,\Delta a &=& +0.0896 & 0.0000
\end{array}$$

From R, B and A we compute

$$\Delta a = +0.0381, \quad \Delta b = -0.0501, \quad \text{and} \quad \Delta r = -0.01465$$

The checks are found by substitution of these in the several equations. The 4 equations

for the observations give the following adjusted values of Δn_i

$$\Delta n_1 = +0{\cdot}0234,\ \Delta n_2 = -0{\cdot}0310,\ \Delta n_3 = -0{\cdot}0770, \text{ and } \Delta n_4 = -0{\cdot}0151;$$

the sum of squares $\left[\frac{(o-u)^2}{\lambda_2}\right]$ (here $= (8-7)\lambda_2$) is consequently

$$= (0{\cdot}0234)^2 + (0{\cdot}0319)^2 + (0{\cdot}0235)^2 + (0{\cdot}0151)^2 = 0{\cdot}00235$$

For this, by the equation (108), we get

$$0{\cdot}01010 - 0{\cdot}00271 - 0{\cdot}00356 - 0{\cdot}00147 = 0{\cdot}00236$$

as the final check of the adjustment

The 4 equations for Δl_i give us

$$\Delta V_1 = +17'2,\ \Delta V_2 = +20'8,\ \Delta V_3 = -2'9, \text{ and } \Delta V_4 = -21'6$$

Thus, by addition of the found corrections to the approximate values,

$$r = 0{\cdot}98535,\quad a = 10{\cdot}0381,\quad b = 9{\cdot}0499,$$
$$V_1 = 0°17'2,\ V_2 = 58°28'6,\ V_3 = 134°57'1, \text{ and } V_4 = 216°30'6,$$

we have the whole system of elements for the next approximation, if they are not the definitive values. In both cases we must compute by them the adjusted values of the co-ordinates, *according to the exact formula*, the resulting differences, obs.—comp., are

Point	Δx	Δy	Δn	Δl
1	$-0{\cdot}0232$	$+0{\cdot}0002$	$-0{\cdot}0232$	$+0{\cdot}0002$
2	$+0{\cdot}0191$	$+0{\cdot}0257$	$+0{\cdot}0320$	$0{\cdot}0000$
3	$+0{\cdot}0166$	$-0{\cdot}0168$	$-0{\cdot}0234$	$-0{\cdot}0001$
4	$-0{\cdot}0123$	$-0{\cdot}0090$	$+0{\cdot}0152$	$0{\cdot}0000$

The sum of the squares, $[(\Delta x)^2 + (\Delta y)^2] = 0{\cdot}00236$, agrees with the above value, which indicates that the approximation of this first hypothesis may have been sufficient. Indeed, the students who will try the next approximation by means of our final differences, will, in this case, find only small corrections.

From the equations A, B, and R, which express the free elements by the original bound elements, Δa, Δb, Δr, we easily compute the equations for the inverse transformation

$$\begin{aligned}\Delta a &= 0{\cdot}4257\ A\\ \Delta b &= -0{\cdot}1444\ A + 0{\cdot}7726\ B\\ \Delta r &= 0{\cdot}0228\ A - 0{\cdot}1752\cdot B + 0{\cdot}25\cdot R\end{aligned}$$

By these, any function of the elements for a given parameter can be expressed as a linear function of the free functions A, B, and R; and by $\lambda_2(A) = 2{\cdot}3490\,\lambda_2$, $\lambda_2(B) = 1{\cdot}2940\,\lambda_2$,

and $\lambda_2(R) = 4\lambda_2$, the mean error is easily found Thus the squares of the mean errors of the co-ordinates x and y are

$$\lambda_2(x) = \left\{2.8400(\ \ 0.4257 + 0.0228\cos V)^2 + 1.2049(-0.1752\cos V)^2 \qquad + 4(0.25\cos V)^2\right\}\lambda_2$$

$$\lambda_2(y) = \left\{2.3490(-0.1444 + 0.0228\sin V)^2 + 1.2043(\ \ 0.7720 - 0.1752\sin V)^2 + 4(0.25\sin V)^2\right\}\lambda_2$$

Only the value $\lambda_2 = 0.00236$, found by the summary criticism, is here very uncertain

XIII SPECIAL AUXILIARY METHODS

§ 57 We have often occasion to use the method of least squares, particularly adjustment by elements, and this sometimes requires so much work that we must try to shorten it as much as possible, even by means which are not quite lawful Several temptations lie near enough to tempt the many who are soon tired by a somewhat lengthened computation, but not so much by looking for subtleties and short cuts And as, moreover, the method was formerly considered the best solution — among other more or less good — not the only one that was justified under the given supposition, it is no wonder that it has come to be used in many modifications which must be regarded as unsafe or wrong After what we have seen of the difference between free and bound functions, it will be understood that *the consequences of transgressions against the method of least squares* stand out much more clearly in the mean errors of the results than in their adjusted values And as — to some extent justly — more importance is attached to getting tolerably correct values computed for the elements, than to getting a correct idea of the uncertainty, *the lax morals with respect to adjustments have taken the form of an assertion to the effect that we can, within this domain, do almost as we like, without any great harm, especially if we take care that a sum of squares, either the correct one or another, becomes a minimum* This, of course, is wrong In a text-book we should do more harm than good by stating all the artifices which even experienced computors have allowed themselves to employ, under special circumstances and in face of particularly great difficulties Only a few auxiliary methods will be mentioned here, which are either quite correct or nearly so, when simple caution is observed

§ 58 When methodic adjustment was first employed, large numbers of figures were used in the computations (logarithms with 7 decimal places), and people often complained of the great labour this caused, but it was regarded as an unavoidable evil, when the elements were to be determined with tolerable exactness We can very often manage, however, to get on by means of a much simpler apparatus, if we do not seek something

which cannot be determined During the adjustment properly so called, we ought to be able to work with three figures But this ideal presupposes that two conditions are satisfied the elements we seek must be small and free of one another, or nearly so; and in both respects it can be difficult enough to protect oneself in time by appropriate transformation Often it is only through the adjustment itself that we learn to know the artifices which would have made the work easy This applies particularly to the mutual freedom of the elements The condition of their smallness is satisfied, if we everywhere use the same preparatory computation as is necessary when the theory is not of linear form

By such means as are used in the exact mathematics, or by a provisional, more or less allowable adjustment, we get, corresponding to the several observations $o_1 \ldots o_n$, a set of values $v_1 \ldots v_n$ which are computed by means of the values $x_0 \ldots z_0$ of the several elements $x \ldots z$, and which, while they satisfy all the conditions of the theory with perfect or at any rate considerable exactness, nowhere show any great deviation from the corresponding observed value It is then these deviations $o_i - v_i$ and $x - x_0 \ldots$ which are made the object of the adjustment, instead of the observations and elements themselves with which, we know, they have mean error in common When in a non-linear theory the equations between the adjusted observation and the elements are of the general form

$$u_i = F(x, \ldots z),$$

they are changed into

$$u_i - v_i = \left(\frac{dF}{dx}\right)_0 (x - x_0) + \ldots + \left(\frac{dF}{dz}\right)_0 (z - z_0) \tag{109}$$

by means of the terms of the first degree in Taylor's series, or by some other method of approximation. If the equations are linear

$$u_i = p_i x + \ldots + r_i z,$$

we have, without any change, for the deviations

$$u_i - v_i = p_i(x - x_0) + \ldots + r_i(z - z_0) \tag{110}$$

No special luck is necessary to find sets of values, v_i, $x_0, \ldots z_0$, whose deviations $o_i - v_i$ show only two significant figures, and then computation by 3 figures is, as far as that goes, sufficient for the needs of the adjustment.

The method certainly requires a considerable extra-work in the preparatory computation, and it must not be overlooked that computations with an exactness of many decimal places will often be necessary in this part; especially v_i ought to be computed with the utmost care as a function of $x_0 \ldots z_0$, lest any uncertainty in this computation should increase the mean errors, so that we dare not put $\lambda_2(o-v) = \lambda_2(o)$

This additional work, however, is not quite wasted, even when the theory is linear The list of the deviations $o_i - v_i$ will, by easy estimates, graphic construction, or directly

by the eye, with tolerable certainty lead to the discovery of gross errors in the series of observations, slips of the pen, etc., which must not be allowed to get into the adjustment. The preliminary rejection of such observations may save a whole adjustment, the ultimate rejection, however, falls under the criticism after the adjustment.

In computing the adjusted values, particularly u_i, after the solution of the normal equations, we ought not to rely too confidently on the transformation of the equations into linear form or into equations of deviations for $o_i - v_i$. Where it is possible, the actual equations $u_i = F(x_i \ldots s)$ ought to be employed, and with the same degree of accuracy as in the computation of v_i. In this way only can we see whether the approximate system of elements and values has been so near to the final result as to justify the rejection of the higher terms in Taylor's series. If not, the adjustment may only be regarded as provisional, and must be repeated until the values of u_i, got by direct computation, agree with the values through $u_i - v_i$ in the linear equations of adjustment.

On the whole the adjustment ought to be repeated frequently till we get a sufficient approximation. This, for instance, is the rule where the observations represent probabilities, for which $\lambda_2(o_i)$ is generally known only as functions of the unknown quantities which the adjustment itself is to give us.

§ 59. The form of the theory, and in particular the selection of its system of elements, is as a rule determined by purely mathematical considerations as to the elegance of the formulæ, and only exceptionally by that freedom between the elements which is wanted for the adjustment. On the other hand it will generally be impossible to arrange the adjustment in such a way that the free elements with which it ends, can all be of direct, theoretical interest. A middle course, however, is always desirable, for the reasons mentioned in the foregoing paragraph, and very frequently it is also possible, if only the theory pays so much respect to the adjustments that it avoids setting up, in the same system, elements between which we may expect beforehand that strong bonds will exist. Thus, in systems of elements of the orbits of planets, the length of the nodes and the distance of the perihelion from the node ought not both to be introduced as elements; *for a positive change in the former will, in consequence of the frequent, small angles of* inclination, nearly always entail an almost equally large negative change in the latter. If a theory says that the observation is a linear function of a single parameter, t, the formula ought not to be written $u = p + qt$, unless all the t's are small, some positive, and others negative, but $u = r + q(t - t_0)$, where t_0 is an average of the parameters corresponding to the observations. If we succeed, in this way, in avoiding all strongly operating bonds, and this can be known by the coefficients of all the normal equations outside the diagonal line becoming numerically small in comparison with the mean proportional between the two corresponding coefficients in the diagonal line, then we have at any rate attained so

much that we need not use in the calculations for the adjustment many more decimal places than about the 3, which will always be sufficient when the elements are originally mutually free, and not during the adjustment are first to be transformed into freedom with painful accuracy in the transformation operations

If, by careful selection of the elements, we even get so far that no sum of the products $[pq]$[1]) in numerical value exceeds about $\frac{1}{50}$ of the mean proportional between the corresponding sums of squares $\sqrt{[pp][qq]}$, or in many cases only $\frac{1}{10}$ of these amounts, then we may consider the bonds between the elements insignificant The normal equations *themselves may then be used to determine the law of error for the elements, we compute* provisionally a first approximation by putting all the small sums of products $= 0$, and in *the second approximation we correct the $[po]$'s by substituting the sums of the products* and the values of the elements as found in the first approximation For instance:

$$[po] - [pq]\,y_1 - \quad - [pr]\,z_1 = [pp]\,x_2 \tag{111}$$

while

$$\lambda_2(x_2) = \frac{1}{[pp]^2}\left\{[pp] + \frac{[pq]^2}{[qq]} + \quad + \frac{[pr]^2}{[rr]}\right\} = \tag{112}$$

$$= 1 \quad \left\{[pp] - \frac{[pq]^2}{[qq]} - \quad - \frac{[pr]^2}{[rr]}\right\} \tag{113}$$

As the errors in these determinations are of the second order, it will not, if the o's themselves are small deviations from a provisional computation, be necessary to make any further approximations.

Even if the bonds between the elements, which are stated in terms of the sums *of the products, are stronger, we can sometimes get them untied without any transformation* If we can get new observations, which are just such functions of the elements that the sums of the products will vanish if they are also taken into consideration, we will of course put off the adjustment until, by introducing them into it, we cannot only facilitate the computation but also increase the theoretical value and clearness of the result And if we can attain freedom of the elements by rejecting from a long series of observations some single ones, we do not hesitate to use this means; especially as such unused observations may very well be employed in the criticism If, for instance, an arctic expedition has made meteorological observations at some fixed station for a little more than a complete year, we shall not hesitate in the adjustment, by means of periodical functions, to leave out the overlapping observations, or to make use of the means of the double values, giving them the weight of single observations

---

[1]) In what follows we write, for the sake of brevity, $[pq]$ for $\left[\frac{pq}{\lambda}\right]$

§ 60 Though of course the fabrication of observations is, in general, the greatest sin which an applied science can commit, there exists, nevertheless, a rather numerous and important class of cases, in which we both can and ought to use a method which just depends on the fabrication of such observations as might bring about the freedom of the theoretical elements As a warning, however, against misuse I give it a harsh name *the method of fabricated observations*

If, for instance, we consider the problem which has served us as an example in the adjustment, both by correlates and by elements, viz the determination of the abscissae for 4 points whose 6 mutual distances have been measured by equally good, bondfree observations, we can scarcely after the now given indications look at the normal equations,

$$\begin{aligned}
o_{12}+o_{13}+o_{14} &= 3x_1-1x_2-1x_3-1x_4 \\
-o_{12}+o_{23}+o_{24} &= -1x_1+3x_2-1x_3-1x_4 \\
-o_{13}-o_{23}+o_{34} &= -1x_1-1x_2+3x_3-1x_4 \\
-o_{14}-o_{24}-o_{34} &= -1x_1-1x_2-1x_3+3x_4,
\end{aligned}$$

without immediately feeling the want of a further observation

$$O = 1x_1+1x_2+1x_3+1x_4,$$

which, if we imagine it to have the same weight $= 1$ as each of the measurements of distance $\lambda_1(o_{rs}) = x_r - x_s$, will give by addition to the others, but without specifying the value of O,

$$\begin{aligned}
O+o_{12}+o_{13}+o_{14} &= 4x_1 \\
O-o_{12}+o_{23}+o_{24} &= 4x_2 \\
O-o_{13}-o_{23}+o_{34} &= 4x_3 \\
O-o_{14}-o_{24}-o_{34} &= 4x_4,
\end{aligned}$$

and consequently determine all 4 abscissae as mutually free and with fourfold weight

What in this and other cases entitles us to fabricate observations is *under-minuteness* in the original problem of adjustment — here, the impossibility of determining any of the abscissae by means of the distances between the points, When we treat such problems in exact mathematics we get simpler, more symmetrical, and easier solutions by introducing values which can only be determined arbitrarily, and so it is also in the theory of observation. But the arbitrariness gets here a greater extent, because not only mean values, but also mean errors must be introduced for greater convenience And while we can always make use of a fabricated observation in indeterminate problems for the complete or partial liberation of the elements, we must here carefully demonstrate, by criticism in each case, that the fabrication we have used has not changed anything which was really determined without it

In the above example, this is seen in the first place by O disappearing from all the adjusted values for the distances $x_r - x_s$, and then by O's own adjusted value, determined as the sum $x_1 + x_2 + x_3 + x_4$, and leading only to the identity $O = O$. The adjustment will consequently neither determine O nor let it get any influence on the other determinations. The mean errors show the same and, moreover, in such a way that the criterion becomes independent of whether O has been brought into the computation as an indeterminate number or with an arbitrary value, for, after the adjustment as well as before, we have for O, $\lambda_1(O) = 1$. *The scale for O is consequently* $= 0$, and this is also generally a sufficient proof of our right to use the method of fabricated observations.

§ 61. *The method of partial eliminations.* When the number of elements is large, it becomes a very considerable task to transform the normal equations and eliminate the elements. The difficulty is nearly proportional to the square of that number. Long before the elements would become so numerous that adjustment by correlates could be indicated, a correct adjustment by elements can become practically impossible. The special criticism is quite out of the question, the summary criticism can scarcely be suggested, and the very elimination must be made easier at any price. If it then happens that some of the elements enter into the expressions for some of the observations only, and not at all in the others, then there can be no doubt that the expedient which ought first to be employed is the partial elimination (before we form the normal equations) of such elements from the observations concerning them. These observations will by this means be replaced by certain functions of two observations or more, *which will generally be bound*, and they will be so in a higher and more dangerous degree the fewer elements we have eliminated. By this proceeding we may, consequently, imperil the whole ensuing adjustment, the foundation of which, we know, is *unbound or free* observations as functions of its elements.

If now it must be granted that the difficulties can become so great that we cannot insist on *an absolute prohibition against illegitimate elimination*, we must on the other hand emphatically warn against every elimination which is not performed through free functions, and much the more so, as it is quite possible, in a great many cases in which abuses have taken place, to remain within the strictly legitimate limits of the free functions, by the use of *"the method of partial eliminations"*.

This is connected with the cases, in which some of the observations, for instance $o_1 \ldots o_m$, according to the theory, depend on certain elements, for instance $x, \ldots y$, which do not occur in the theoretical expression for any other of the observations. Our object is then, by the formation of the normal equations to separate $o_1 \ldots o_m$ as a special series of observations. We begin by forming the partial normal equations for this, and then immediately perform the elimination of $x, \ldots y$ from them, without taking into consideration whether these equations alone would be sufficient for a determination of the other elements.

As soon as $x \ldots y$ are eliminated, the process of elimination is suspended. The transformed equations containing these elements (which now represent functions that are free of all observations, and functions which depend only on the remaining elements $z, \ldots u$), are put aside till we come back to the determination of $x \ldots y$. The other partially transformed normal equations, originating in the group $o_1 \ldots o_m$, are on the other hand to be added, term by term, to the normal equations for the elements $z, \ldots u$, formed out of the remaining observations, before the process of elimination is continued for these elements.

That this proceeding is quite legitimate becomes evident if we imagine the elements $x \ldots y$ transformed into the elements $x' \ldots y'$, which are free of $z \ldots u$, and then imagine $x' \ldots y'$ inserted instead of $x \ldots y$ in the original equations for the observations. For then all the sums of products with the coefficients of $x' \ldots y'$ will identically become $= 0$, and the sums of squares and sums of products for the separated part of the observations will, as addenda in the coefficients of the normal equations (compare (57)), come out, immediately, with the same values as now the transformed normal equations.

As an example we may treat the following series of measurements of the position of 3 points on a straight line. The mode of observation is as follows. We apply a millimeter scale several times along the straight line and then each time read off by inspection with the unaided eye either the places of all the points against the scale or the places of two of them. The readings for each point are found in its separate column, and those on the same row belong to the same position of the scale. (Considered as absolute abscissa-observations such observations are bound by the position of the zero by every laying down of the scale, but those bonds are evidently loosened by our taking up the position against the scale of an arbitrarily selected fixed origin y_r as an element beside the abscissae x_1, x_2, x_3 of the three points.) All mean errors are supposed to be equal.

Position of the Scale	Point I	Point II	Point III	Eliminated free Elements	
	mm.	mm.	mm.	mm.	
1	6·9	27·54		$17·22 = y_1 + \frac{1}{2}(x_1 + x_2)$	Weight = 2
2	8·35		54·95	$31·65 = y_2 + \frac{1}{2}(x_1 + x_3)$	
3	7·9		54·5	$31·20 = y_3 + \frac{1}{2}(x_1 + x_3)$	
4		21·16	47·2	$34·18 = y_4 + \frac{1}{2}(x_2 + x_3)$	
5		10·74	36·7	$23·72 = y_5 + \frac{1}{2}(x_2 + x_3)$	
6		4·06	30·1	$17·08 = y_6 + \frac{1}{2}(x_2 + x_3)$	
7	31·45	51·98	78·06	$53·83 = y_7 + \frac{1}{3}(x_1 + x_2 + x_3)$	Weight = 3
8	32·9	56·5	79·5	$56·30 = y_8 + \frac{1}{3}(x_1 + x_2 + x_3)$	
9	9·6	30·3	56·22	$32·04 = y_9 + \frac{1}{3}(x_1 + x_2 + x_3)$	
10	20·16	40·76	66·8	$42·58 = y_{10} + \frac{1}{3}(x_1 + x_2 + x_3)$	
11	18·9	39·5	65·56	$41·32 = y_{11} + \frac{1}{3}(x_1 + x_2 + x_3)$	

As the theoretical equation for the i^{th} observation in the s^{th} column has the form

$$o_{i,s} = y_i + x_s,$$

and every observation, therefore, is a function of only two elements, there is every reason to use the method of partial elimination If we choose first to eliminate the ys, we have consequently to form normal equations for each of the 11 rows Where only two points are observed these normal equations get the form

$$\begin{aligned} o_r + o_s &= 2y_i + x_r + x_s \\ o_r &= y_i + x_r \\ o_s &= y_i \qquad + x_s, \end{aligned}$$

for three points the form of the normal equations is

$$\begin{aligned} o_1 + o_2 + o_3 &= 3y_i + x_1 + x_2 + x_3 \\ o_1 &= y_i + x_1 \\ o_2 &= y_i \qquad + x_2 \\ o_3 &= y_i \qquad\qquad + x_3 \end{aligned}$$

Of these equations those referring to the y_i have given the eliminated free elements stated above to the right of the observations after the perpendicular

By subtracting these equations from the corresponding other equations we get, in the cases where there are 2 points

$$\begin{aligned} o_r - \tfrac{1}{2}(o_r + o_s) &= \tfrac{1}{2}x_r - \tfrac{1}{2}x_s \\ o_s - \tfrac{1}{2}(o_r + o_s) &= -\tfrac{1}{2}x_r + \tfrac{1}{2}x_s, \end{aligned}$$

and in cases where there are 3 points

$$\begin{aligned} o_1 - \tfrac{1}{3}(o_1 + o_2 + o_3) &= \tfrac{2}{3}x_1 - \tfrac{1}{3}x_2 - \tfrac{1}{3}x_3 \\ o_2 - \tfrac{1}{3}(o_1 + o_2 + o_3) &= -\tfrac{1}{3}x_1 + \tfrac{2}{3}x_2 - \tfrac{1}{3}x_3 \\ o_3 - \tfrac{1}{3}(o_1 + o_2 + o_3) &= -\tfrac{1}{3}x_1 - \tfrac{1}{3}x_2 + \tfrac{2}{3}x_3 \end{aligned}$$

By forming the sum of these differences for each column, and counting, on the right side of the equations, how often each element occurs with one other or with two others, we consequently get the ultimate normal equations

$$\begin{aligned} -168.98 &= \tfrac{23}{6}x_1 - \tfrac{13}{6}x_2 - \tfrac{10}{6}x_3 \\ -37.71 &= -\tfrac{13}{6}x_1 + \tfrac{23}{6}x_2 - \tfrac{10}{6}x_3 \\ +206.69 &= -\tfrac{10}{6}x_1 - \tfrac{10}{6}x_2 + \tfrac{20}{6}x_3 \end{aligned}$$

The case is here simple enough to be solved by a fabricated observation How is its most advantageous form found, when its existence is given?

Answer $\dfrac{759\,o}{28712} = \dfrac{x_1}{114} + \dfrac{x_2}{96} + \dfrac{x_3}{79}$, weight $= 28712$,

after which we get the normal equations

$$\frac{180}{114}o - 168{\cdot}98 = \frac{759}{114}x_1$$
$$\frac{759}{84}o - 37{\cdot}71 = \frac{180}{84}x_2$$
$$\frac{759}{78}o + 200{\cdot}60 = \frac{180}{78}x_3\,,$$

consequently,

$$x_1 = o - \overset{mm}{25}{\cdot}38\,,\quad x_2 = o - \overset{mm}{4}{\cdot}77\,,\ \text{and}\ x_3 = o + \overset{mm}{21}{\cdot}24$$

From these we now compute the y's

$$\begin{array}{ll} y_1 = 32{\cdot}205 - o\,, & y_7 = 56{\cdot}80 - o\,, \\ y_2 = 33{\cdot}72 - o\,, & y_8 = 58{\cdot}27 - o\,, \\ y_3 = 33{\cdot}27 - o\,, & y_9 = 85{\cdot}01 - o\,, \\ y_4 = 25{\cdot}945 - o\,, & y_{10} = 45{\cdot}55 - o\,, \\ y_5 = 15{\cdot}485 - o\,, & y_{11} = 44{\cdot}20 - o\,. \\ y_6 = 8{\cdot}845 - o\,, & \end{array}$$

We need not here state the adjusted values for the several observations, nor their differences, of which it is enough to say that their sum vanishes both for each row and for each column; their squares on the other hand, will be found to be:

	I	II	III	Total.	
	0002	0002		0004	
	1		0001	2	
	1		1	2	$\Sigma = 0028$
		2	2	4	
		6	6	12	
		2	2	4	
	0	25	4	38	
	1	0	1	2	
	9	36	0	54	$\Sigma = \cdot0110$
	1	0	1	2	
	1	4	0	14	
Total	0025	0077	0036	0138	

For the summary criticism we notice that the number of observations is 27, the number of the elements is $3 + 11 - 1 = 13$, divisor consequently $= 14$ (one element being wholly engaged by the fabricated observation o). The unit of the mean error is therefore determined by $E^2 = 0{\cdot}0010$, and the mean error on single reading $\pm\overset{mm}{0}{\cdot}032$, which agrees well with what we may expect to attain by practice in estimates of tenth parts.

As to special criticism it is here, where the weights of the eliminated free functions are respectively 2 and 3 times the weight of the single observation, while the weights of x_1, x_2, and x_3 after the adjustment become respectively $\frac{760}{114}$, $\frac{760}{84}$, and $\frac{760}{71}$, very easy to compute the scales

$$1 - \frac{\lambda_2(u)}{\lambda_2(o)} = 1 - \frac{1}{\text{Weight after the adjustment}}$$

With 760 as common denominator we find for the several scales and the sums of their most natural groups

	I	*II*	*III*		
1	327	327		654	
2	331 5		331 5	663	
3	331 5		331 5	663	$\Sigma = 3996$
4		336	336	672	
5		336	336	672	
6		336	336	672	
7	436	442	448	1326	
8	436	442	448	1326	
9	436	442	448	1326	$\Sigma = 6630$
10	436	442	448	1326	
11	436	442	448	1326	
	3170	3545	3911	10626	

The comparison with the sums of squares in the groups, divided by E^2, shows then for point I 2 5 instead of $\frac{3170}{760} = 4\,2 \pm \sqrt{8\,4}$, for point II 7 7 instead of $4\,7 \pm \sqrt{9{\cdot}4}$, for point III 8 6 instead of $5\,1 \pm \sqrt{10\,2}$, for all positions of the scale with two readings 2 8 instead of $5\,3 \pm \sqrt{10\,6}$, and for positions with 3 readings 11 0 instead of $8\,7 \pm \sqrt{17\,4}$ The limit of the mean error is consequently reached only in the group of point II, where $(7\,7 - 4{\cdot}7)^2 = 9{\cdot}0 < 9{\cdot}4$, and it is nowhere exceeded We have a check by summing the scales,

$$\frac{10626}{760} = 14 = 27 - 11 - 3 + 1$$

---

§ 62 In such cases in which the circumstances and weights of the observations are distributed in some regular way, this will often facilitate the treatment of the normal equations *The elimination of the elements and the transformation of the normal equations into such whose left hand sides can be regarded as unbound observations* as they are free

functions of the original observations, need not always be so firmly connected with one another as in the ordinary method If we, in a suitable way, take advantage of regularity in the observations, and thereby are able to find a transformation which sets the normal equations free, then the determination of the several elements will scarcely throw any material obstacles in our way But in order to find out any special transformations, we must know the general form of the changes of the normal equations resulting from transformation of the original elements into such as are any homogeneous linear funtions of them whatever

If the equations for the unbound observations in terms of the original elements have been

$$o_i = p_i x + q_i y + r_i z,$$

the normal equations will be

$$\begin{aligned} [po] &= [pp]x + [pq]y + [pr]z \\ [qo] &= [qp]x + [qq]y + [qr]z \\ [ro] &= [rp]x + [rq]y + [rr]z \end{aligned}$$

And if we wish to substitute new elements, ξ, η, and ζ, for the old ones, we make use of substitutions in which *the original elements are represented as functions of the new ones*, therefore

$$\left.\begin{aligned} x &= h_1\xi + k_1\eta + l_1\zeta \\ y &= h_2\xi + k_2\eta + l_2\zeta \\ z &= h_3\xi + k_3\eta + l_3\zeta \end{aligned}\right\} \qquad (114)$$

The equations for the observations then have the form

$$o_i = (p_i h_1 + q_i h_2 + r_i h_3)\xi + (p_i k_1 + q_i k_2 + r_i k_3)\eta + (p_i l_1 + q_i l_2 + r_i l_3)\zeta \qquad (115)$$

The new normal equations may be formed from these, but the form becomes very cumbrous, the equation which specially refers to ξ being

$$[(ph_1+qh_2+rh_3)\,o] = [(ph_1+qh_2+rh_3)^2]\,\xi + [(ph_1+qh_2+rh_3)(pk_1+qk_2+rk_3)]\,\eta + \\ + [(ph_1+qh_2+rh_3)(pl_1+ql_2+rl_3)]\,\zeta$$

The computation ought not to be performed according to the expressions for the coefficients which come out when we get rid of the round brackets under the signs of summation [ ]. But it is easy to give the rule of the computation with full clearness The old normal equations are first treated exactly as if they were equations for unbound observations, for x, y, and z, respectively, expressed by the new elements, consequently by multiplication, by *columns*, by h_1, h_2, and h_3 and addition, by multiplication by k_1, k_2, and k_3 and addition, and by multiplication by l_1, l_2, and l_3 and succeeding addition Thereby, certainly, we get the new normal equations, but still with preservation of the old elements

$$\left.\begin{aligned}
[(ph_1+qh_2+rh_3)o] &= [(ph_1+qh_2+rh_3)p]x+[(ph_1+qh_2+rh_3)q]y+[(ph_1+qh_2+rh_3)r]z\\
[(pk_1+qk_2+rk_3)o] &= [(pk_1+qk_2+rk_3)p]x+[(pk_1+qk_2+rk_3)q]y+[(pk_1+qk_2+rk_3)r]z\\
[(pl_1+ql_2+rl_3)o] &= [(pl_1+ql_2+rl_3)p]x+[(pl_1+ql_2+rl_3)q]y+[(pl_1+ql_2+rl_3)r]z
\end{aligned}\right\}(116)$$

The second part of the operation must therefore consist in the substitution of the new elements for the original ones in the right hand sides of these equations. In order to find the coefficients of ξ, η, and ζ, we must therefore here again multiply the sums of the products, *now by rows*, by

$$\begin{matrix} h_1, & h_2, & h_3 \\ k_1, & k_2, & k_3 \\ l_1, & l_2, & l_3 \end{matrix}$$

and add them up.

Example. It happens pretty often, for instance in investigations of scales for linear measures, that there is symmetry between the elements, two and two, x_r and x_{n-r}, so that for instance the normal equation which specially refers to x_r, has the same coefficients, only in inverted order, as the normal equation corresponding to x_{n-r}, of course, irrespective of the two observed terms $[po]$ on the left hand sides of the equations. Already P. A. Hansen pointed out that this indicates a transformation of the elements into the mean values $s_r = \frac{1}{2}(x_r+x_{n-r})$ and their half differences $d_r = \frac{1}{2}(x_r-x_{n-r})$. In this case therefore the equations for the old elements by the new ones have the form

$$\begin{aligned} x_r &= s_r + d_r \\ x_{n-r} &= s_r - d_r, \end{aligned}$$

and the transformation of the normal equations is, consequently, performed just by forming sums and differences of the original coefficients. If the normal equations are

$$\begin{aligned}
[ao] &= 4x+3y+2z+1u\\
[bo] &= 3x+6y+4z+2u\\
[co] &= 2x+4y+6z+3u\\
[do] &= 1x+2y+3z+4u,
\end{aligned}$$

the procedure is as follows:

$$[ao]+[do] = 5x+5y+5z+5u = 10\frac{x+u}{2}+10\frac{y+z}{2}$$

$$[bo]+[co] = 5x+10y+10z+5u = 10\frac{x+u}{2}+20\frac{y+z}{2}$$

$$[ao]-[do] = 3x+1y-1z-3u = 6\frac{x-u}{2}+2\frac{y-z}{2}$$

$$[bo]-[co] = 1x+2y-2z-1u = 2\frac{x-u}{2}+4\frac{y-z}{2}$$

As in this example, we always succeed in separating the mean values from the half differences, as two mutually free systems of functions of the observations

§ 63 The great simplification that results when the observations are mere repetitions, in contradistinction to the general case when there are varying circumstances in the observations, is owing to the fact that the whole adjustment is then reduced to the determination of the mean values and the mean errors of the observations Before an adjustment, therefore, we not only take the means of any observations, which are strictly speaking repetitions, but we also save a good deal of work in the cases which only approximate to repetitions, viz those where the variations of circumstances have been small enough to allow us to neglect their products and squares It has not been necessary to await the systematic development of the theory of observations to know how to act in such cases

When astronomers have observed the place of a planet or a comet several times in the same night, they form a mean time of observation t, a mean right ascension a, and a mean declination δ, and consider a and δ the spherical co-ordinates of the star at the time t

With the obvious extensions this is what is called the *normal place* method, the most important device in practical adjustment Such observations whose essential circumstances have "small" variations, are, before the adjustment, brought into a normal place, by forming mean values both for the observed values themselves and for each of their essential circumstances, and on the supposition that the law which connects the observations and circumstances, holds good also, without any change, with respect to their mean values

Much trouble may be spared by employing the normal place method The question is, whether we lose thereby in exactness, and then how much.

We shall first consider the case where the unbound observations o are linear functions of the varying essential circumstances $x, \ldots z$, the equation for the observations being

$$\lambda_1(o) = a + bx + \ldots + dz$$

With the weights v we form the normal equations:

$$[vo\,] = a[v\,] + b[vx] + \ldots + d[vz] \qquad (117)$$

$$\left.\begin{array}{l} [vxo] = a[vx] + b[vx^2] + \ldots + d[vxz] \\ \quad\vdots \\ [vzo] = a[vz] + b[vzx] + \ldots + d[vz^2]. \end{array}\right\} \qquad (118)$$

If the whole series of observations is gathered into a single normal place, O, corresponding to the circumstances $X \ldots Z$, and with the weight V, we shall have:

$$
\begin{aligned}
V &= [v] \\
VO &= [vo] \\
VX &= [vx] \\
&\;\;\vdots \\
VZ &= [vz],
\end{aligned}
$$

and as

$$O = a + bX + \dots + dZ, \tag{117a}$$

this normal place will exhaust the normal equation (117) corresponding to the constant term, both with respect to mean value and mean error. But if we make the other normal *equations free of (117), we get, by the correct method of least squares,*

$$
\left.
\begin{aligned}
[v(o-O)(x-X)] &= b[v(x-X)^2] + \dots + d[v(x-X)(z-Z)] \\
&\;\;\vdots \\
[v(o-O)(z-Z)] &= b[v(x-X)(z-Z)] + \dots + d[v(z-Z)^2]
\end{aligned}
\right\} \tag{118a}
$$

for the determination of the elements $b \dots d$, and these determinations are lost completely *if the whole series is gathered into a single normal place. Certainly, the coefficients of these* equations (118a) are small quantities of the second order, if the $x - X$ and $z - Z$ are small of the first order.

If, on the other hand, we split up the series, forming for each part a normal place, and adjusting these normal places instead of the observations according to the method of the least squares, then the normal equation corresponding to the constant term is still exhausted by the normal place method, and besides this determination of $a + bX + \dots + dZ$ the normal place method now also affords a determination of the other elements $b \dots d$, in such a way, however, that we suffer a loss of the weights for their determination. This loss can become great, nay total, if the normal places are selected in a way that does not suit the purpose; but it can be made rather insignificant by a suitable selection of normal places in not too small a number.

Let us suppose, in order to simplify matters, that the observations have only one variable essential circumstance x, of which their mean values are linear functions, consequently

$$\lambda_1(o) = a + bx,$$

and that the x's are uniformly distributed within the utmost limits, x_0 and x_1, we then let each *normal place encompass an equally large part of this interval,* and we shall find then, this being the most favourable case, with n normal places, that the weight on the adjusted value of the element b becomes $1 - \left(\frac{1}{n}\right)^2$, if by a correct adjustment by elements the corresponding *weight is taken as unity. The loss is thus, at any rate, not very great. And it can be* made still smaller, if the distribution of the essential circumstance of the observations is

uneven, and if we can get a normal place everywhere where the observations become particularly frequent, while empty spaces separate the normal places from each other

The case is analogous also when the observations are still functions of a single or a few essential circumstances, but the function is of a higher degree, or transcendental For it is possible also to form normal places in these cases, and we can do so not only when the variations of the circumstances can be directly treated as infinitely small within each normal place, which case by Taylor's theorem falls within the given rule. For if we have at our disposal a provisional approximate formula, $y = f(x)$, and have calculated the deviation from this, $o - y$, of every observation (considering the deviations as observations with the essential circumstances and mean errors of the original observations), then we can use mean numbers of deviations for reciprocally adjacent circumstances as corrections which, added to the corresponding values from the approximate formula, give the normal values Further, it is required here only that no normal place is made so comprehensive that the deviations within its limits do not remain linear functions of the essential circumstances

Also here part of the correctness is lost, and it is difficult to say how much The loss is, under equal circumstances, smaller, the more normal places we form With twice (or three times) as many normal places as the number of the unknown elements of the problem, it will rarely become perceptible With due regard to the essential circumstances and the distribution of the weights we can reduce it, using empty spaces as boundaries between the normal places

A suitable distribution of the normal places also depends on what function the observations are of their essential circumstances As to this, however, it is, as a rule, sufficient to know the behaviour of the integral algebraic functions, as we generally, when we have to do with functions which are essentially different from these, will try through transformations of the variables to get back to them and to certain functions which resemble them in this respect

We need only consider the cases in which we have only one variable essential circumstance, of which the mean value of the observation is an algebraic function of the r^{th} degree We are able then, on any supposition as to the distribution of the observations, o, and their essential circumstances, x, and weights, v, to determine $r+1$ substitutive observations, O, together with the essential circumstances, X, and weights, V, belonging to them, in such a way that they treated according to the method of the least squares will give the same results as the larger number of actual observations The conditions are

$$\left.\begin{array}{l} [ov] = O_1 V_1 + \dots + O_r V_r \\ \vdots \\ [ox^r v] = X_1^r O_1 V_1 + \dots + X_r^r O_r V_r \end{array}\right\} \quad (119)$$

and

$$\left.\begin{array}{l} [v] \quad = V_0 + \dots + V_r \\ \dots \\ [x^{2r} v] = X_0^{2r} V_0 + \dots + X_r^{2r} V_r \end{array}\right\} \qquad (120)$$

These $3r+2$ equations are not quite sufficient for the determination of the $3r+3$ unknowns. We remove the difficulty in the best way by adding the equation

$$[x^{2r+1} v] = X_0^{2r+1} V_0 + \dots + X_r^{2r+1} V_r.$$

The elimination of the V's (and O's) then leads to an equation of the $r+1$ degree, whose roots $X_0, \dots X_r$ are all real quantities, if the given x's have been real and the v's positive. When the roots are found, we can compute, first $V_0, \dots V_r$ and afterwards $O_0, \dots O_r$, by means of two systems of $r+1$ linear equations with $r+1$ unknowns.

If, for instance, the essential circumstances of the actual observations are contained in the interval from -1 to $+1$, and if the observations are so numerous and so equally distributed that they may be looked upon as continuous with constant mean error everywhere in this interval, if, further, the sum of the weights $= 2$, then the distribution of the substitutive observations will be symmetrical around 0, and, for functions of the lowest degrees, be

$$r = 0 \begin{cases} X = \ \ \cdot 000 \\ V = \ \ 2\cdot 000 \end{cases}$$

$$r = 1 \begin{cases} X = -\cdot 577, \ +\cdot 577 \\ V = \ \ 1\cdot 000, \ \ 1\cdot 000 \end{cases}$$

$$r = 2 \begin{cases} X = -\cdot 775, \ \ \cdot 000, \ +\cdot 775 \\ V = \ \ \cdot 556, \ \ \cdot 889, \ \ \cdot 556 \end{cases}$$

$$r = 3 \begin{cases} X = -\cdot 861, \ -\cdot 340, \ +\cdot 340, \ +\cdot 861 \\ V = \ \ \cdot 348, \ \ \cdot 652, \ \ \cdot 652, \ \ \cdot 348 \end{cases}$$

$$r = 4 \begin{cases} X = -\cdot 906, \ -\cdot 538, \ \ \cdot 000, \ +\cdot 538, \ +\cdot 906 \\ V = \ \ \cdot 237, \ \ \cdot 479, \ \ \cdot 569, \ \ \cdot 479, \ \ \cdot 237 \end{cases}$$

$$r = 5 \begin{cases} X = -\cdot 932, \ -\cdot 661, \ -\cdot 239, \ +\cdot 239, \ +\cdot 661, \ -\cdot 932 \\ V = \ \ \cdot 171, \ \ \cdot 361, \ \ \cdot 468, \ \ \cdot 468, \ \ \cdot 361, \ \ \cdot 171 \end{cases}$$

$$r = 6 \begin{cases} X = -\cdot 949, \ -\cdot 742, \ -\cdot 406, \ \ \cdot 000, \ +\cdot 406, \ +\cdot 742, \ +\cdot 949 \\ V = \ \ \cdot 129, \ \ \cdot 280, \ \ \cdot 382, \ \ \cdot 418, \ \ \cdot 382, \ \ \cdot 280, \ \ \cdot 129 \end{cases}$$

If, in another example, the distribution of the observations is, likewise, continuous, but the weights within the element dx proportional to e^{-x^2}, consequently symmetrical with maximum by $x = 0$, then the distribution for the lowest degrees, the only ones of any practical interest, will be

$$r = 0\begin{cases} X = & 000 \\ V = & 2\,000\,' \end{cases}$$

$$r = 1\begin{cases} X = & -1\,000, & +1\,000 \\ V = & 1\,000, & 1\,000\,' \end{cases}$$

$$r = 2\begin{cases} X = & -1\,732, & 000, & +1\,732 \\ V = & 333, & 1\,333, & 333\,' \end{cases}$$

$$r = 3\begin{cases} X = & -2\,334, & -\ 742, & +\ 742, & +2\,334 \\ V = & 002, & 008, & 008, & 002\,' \end{cases}$$

$$r = 4\begin{cases} X = & -2\,857, & -1\,356, & 000, & +1\,356, & +2\,857 \\ V = & 023, & 444, & 1\,067, & 444, & 023\,' \end{cases}$$

$$r = 5\begin{cases} X = & -3\,324, & -1\,880, & -\ 617, & +\ 617, & +1\,880, & +3\,324 \\ V = & 005, & 177, & 818, & 818, & 177, & 005\,' \end{cases}$$

$$r = 0\begin{cases} X = & -3\,750, & -2\,307, & -1\,154, & 000 & +1\,154, & +2\,307, & 3\,750 \\ V = & 001, & 002, & 480, & 014, & 480, & 062, & 001 \end{cases}$$

If we were able now to represent these substitutive observations as normal places, then we should be able also, by the use of such tables in analogous cases, to prevent any loss of exactness. It would be possible entirely to evade the application of the method of the least squares, we had but to form such qualified normal places in just the same number as the adjustment formula contains elements that are to be determined This, however, is not possible Certainly, we can obtain normal places corresponding to the required values of the essential circumstance but we cannot by a simple formation of mean numbers give them the weight which each of them ought to have, without employing some of the observations twice, others not at all By taking into consideration how much the extreme normal places from this reason must lose in weight, compared to the substitutive observations, we can estimate how many per cent the loss, in the worst case, can amount to In the first of our examples we find the loss to be 0, for $r = 0$ and $r = 1$, but for $r = 2$ we lose 15, for $r = 3$ we lose 10, for $r = 4$ we lose 20, and for greater values of r 21 p c

Example Eighteen unbound observations, equally good, $\lambda_2(o) = \frac{1}{18}$, correspond to an essential circumstance whose values are distributed as the prime numbers p from 1049 to 1171 Taking $(p-1105)$ 100 $= x$ as the essential circumstance of the observation o, we have

x	o	x	o	x	o
− 50	− 41	− 14	− ·15	+ 18	− 24
− 54	+ 50	− 12	− 72	+ 24	+ 09
− 44	− ·03	− 08	+ 38	+ 40	+ 30
− 42	− 15	− 02	− 21	+ 48	+ 12
− 30	+ 48	+ ·04	+ 21	+ 58	− 24
− 18	+ 18	+ 12	+ 40	+ 00	− 90

Dividing these observations into groups indicated by the horizontal lines we get the 6 normal places

x	o	weight
− 550	+ 045	2
− ·407	+ ·100	3
− 108	− ·094	5
+ 145	+ ·115	4
+ 470	+ 255	2
+ 020	− 315	2

If we suppose the mean values of the observations to be a function of the third, eventually second, degree of x, $\lambda_1(o) = a + bx + cx^2 + dx^3$, we have by ordinary application of the adjustment by elements the normal equations

$$
\begin{aligned}
0\,72 &= 216\,00\,a - 1\,20\,b + 29{\cdot}08\,c + 1{\cdot}94\,d \\
-3{\cdot}07 &= -1\,20\,a + 29\,98\,b + 1\,04\,c + 8\,11\,d \\
-1\,08 &= 20\,08\,a + 1\,94\,b + 8{\cdot}11\,c + 1\,21\,d \\
-1\,44 &= 1{\cdot}94\,a + 8\,11\,b + 1\,21\,c + 2{\cdot}50\,d
\end{aligned}
$$

By the free equations

$$
\begin{aligned}
0\,72 &= 210\,00\,a - 1{\cdot}20\,b + 29{\cdot}08\,c + 1{\cdot}94\,d \\
-3\,09 &= 29\,97\,b + 2\,11\,c + 8\,12\,d \\
-1{\cdot}70 &= 8{\cdot}80\,c + {\cdot}27\,d \\
-\ 64 &= {\cdot}305\,d
\end{aligned}
$$

we get

$$
\begin{aligned}
a &= +\ 09, & a' &= +\ 10, \\
b &= +\ 40, & b' &= -\ 07, \\
c &= -\ 80, & c' &= -\ {\cdot}47, \\
d &= -1{\cdot}77,
\end{aligned}
$$

where a', b', c' are the coefficients in the functions of second degree, obtained by presupposing $d = 0$.

Now, by application of the normal places instead of the original observations, we obtain on the same suppositions the normal equations

$$
\begin{aligned}
6.72 &= 216.00\,a - 1.20\,b + 29.45\,c + 1.87\,d \\
-2.84 &= -1.20\,a + 29.45\,b + 1.87\,c + 7.93\,d \\
-.54 &= 29.45\,a + 1.87\,b + 7.03\,c + 1.14\,d \\
-1.57 &= 1.87\,a + 7.93\,b + 1.14\,c + 2.45\,d
\end{aligned}
$$

By the free equations

$$
\begin{aligned}
6.72 &= 216.00\,a - 1.20\,b + 29.45\,c + 1.87\,d \\
-2.80 &= 29.44\,b + 2.03\,c + 7.94\,d \\
-1.26 &= 3.77\,c + .34\,d \\
-.76 &= 2.63\,d,
\end{aligned}
$$

we get

$$
\begin{aligned}
a &= +.07, \quad a' = +.08, \\
b &= +.09, \quad b' = -.07, \\
c &= -.07, \quad c' = -.39, \\
d &= -2.88
\end{aligned}
$$

A comparison between these two calculations, particularly between the leading coefficients in the free equations, shows that the loss of weight amounts to 1 − [illegible], or 14 per cent. But it is only in the equation for d that the loss is so great; in the equations for b and c, respectively, it is only two and one per cent.

Our normal places are very good if the function is only of the first or second degree; for the function of third degree they can be admitted even though the values of the elements a, b, c, d have changed considerably. For functions of 4[th] or higher degrees these normal places would prove insufficient.

§ 64. That *graphical adjustment* is a means which can carry us through great difficulties, we have shown already in practice by applying it to the drawing of curves of errors. The remarkable powers of the eye and the hand must, like a deus ex machina, help us where all other means fail.

Adjustment by drawing is restricted only by one single condition: if we are to represent a relation between quantities by a plane curve, there must be only two quantities; one of these, represented by the ordinate, is, or is considered to be, the observed value, and the other, represented by the abscissa, is considered the only essential circumstance on which the observed value depends.

Examples of graphical adjustment with two essential circumstances do occur, however, for instance in weather-charts. In periodic phenomena polar co-ordinates are preferred. But otherwise each observation is represented by a point whose ordinate and

abscissa are, respectively, the observed value and its essential circumstance, and the adjustment is performed by free-hand drawing of a curve which satisfies the two conditions of being free from irregularities and going as near as possible to the several points of observation The smoothness of the curve in this process plays the part of the theory, and it is a matter of course that we succeed relatively best when the theory is unknown or extremely intricate, when, for instance we must confine ourselves to requiring that the phenomenon must be continuous within the observed region, or be a single valued function But also such a theoretical condition as, for instance, the one that the law of dependence must be of an integral, rational form, may be successfully represented by graphical adjustment, if the operator has had practice in the drawing of parabolas of higher degrees And we have seen that also such functional forms as have the rapid approximation to an asymptote which the curves of error demand, lie within the province of the graphical adjustment

As for the approximation to the several observed points, the idea of the adjustment implies that a perfect identity is not necessary, only, the curve must intersect the ordinates so near the points as is required by the several mean errors or laws of errors If, after all, we know anything as to the exactness of the several observations before we make the adjustment, this ought to be indicated visibly on the drawing-paper and used in the graphical adjustment We cannot pay much regard, of course, to the presupposed typical form and other properties of the law of errors, but something may be attained, particularly with regard to the number of similar deviations

If we know nothing whatever as to the exactness of the several observations, or only that they are all to be considered equally good, there can be only a single point in our figure for each observation. In a graphical adjustment, however, we can and ought to take care that the curve we draw has the same number of observed points on each side of it, not only in its whole extent, but also as far as possible for arbitrary divisions If we know the weights of the observations, they may be indicated on the drawing, and observations with the weight n count n-fold

In contradistinction to this it is worth while to remark that, with the exception only of bonds between observations, represented by different points, it is possible to lay down on the paper of adjustment almost all desirable information about the several laws of errors Around each point whose co-ordinates represent the mean values of an observation and of its essential circumstance, a curve, the curve of mean errors, may be drawn in such a way that a real intersection of it with any curve of adjustment indicates a deviation less than the mean error resulting from the combination of the mean errors of the observed value and that of its essential circumstance, if this is also found by observation, while a passing over or under indicates a deviation exceeding the mean error Evidently, drawings furnished with such indications enable us to make very good adjustments

If the laws of errors both for the observation and for its circumstance are typical, then the curve of mean errors is an ellipse with the observed points in its centre

If, further, there are no bonds between the observation and its circumstance, then the ellipse of mean errors has its axes parallel to the ordinate and the abscissa, and their lengths are double the respective mean errors

If the essential circumstance of the observation, the abscissa, is known to be free of errors, the ellipse of the mean errors is reduced to the two points on the ordinate, distant by the mean error of the observation from the central point of observation In special cases other means of illustrating the laws of errors may be used If, for instance, the mean errors as well as the mean values are continuous functions of the essential circumstance of the observation, continuous curves for the mean errors may be drawn on the adjustment paper

The principal advantages of the graphical adjustment are its indication of gross errors and its independence of a definitely formulated theory By measuring the ordinates of the adjusted curve we can get improved observations corresponding to as many values of the circumstance or abscissa as we wish, and we can select them as we please within the limits of the drawing, But these adjusted observations are strongly bound together, and we have no indication whatever of their mean errors Consequently, no other adjustment can be based immediately upon the results of a graphical adjustment

On the other hand, graphical adjustment can be very advantageously combined with interpolations, both preceding and following, and we shall see later on that by this means we can remedy its defects, particularly its limited accuracy and its tendency to place too much confidence in the observations, and too little in the theory, i e, to give an under-adjustment.

By drawing we attain an exactness of only 3 or 4 significant figures, and that is frequently insufficient The scale of the drawing must be chosen in such a way that the errors of observations are visible, but then the dimensions may easily become so large that no paper can contain the drawing In order to give the eye a full grasp of the figure, the latter must in its whole course show only small deviations from the straight line, which is taken as the axis of abscissae This is a practical hint, founded upon experience The eye can judge of the smoothness of other curves also, but not by far so well as of that of a straight line, And if the line forms a large angle with the axis of the abscissae, then the exactness is lost by the flat intersections with the ordinates Therefore, as a rule, it is not the original observations that are marked on the paper when we make a graphical adjustment, but only their differences from values found by a preceding interpolation

In order to avoid an under-adjustment, we must allow $\frac{1}{3}$ of the deviations of the curve from the observation-points to surpass the mean errors It is further essential that

the said interpolation is based on a minimum number of observed data, and after the graphical adjustment has been made, it is safe to try another interpolation using a smaller number of the adjusted values as the base of a new interpolation and a repeated graphical adjustment

If the results of a graphical adjustment are required only in the form of a table representing the adjusted observations as a function of the circumstance as argument, this table also ought to be based on an interpolation between relatively few measured values, the interpolated values being checked by comparison with the corresponding measured values A table of exclusively measured values will show too irregular differences.

When we have corrected those values by measuring the ordinates in a curve of graphical adjustment, they may be employed instead of the observations as a sort of normal places It has been said, however, and it deserves to be repeated, that they must not be adjusted by means of the method of the least squares, like the normal places properly so called But we can very well use both sorts of normal places, in a *just sufficient number*, for the computation of the unknown elements of the problem, according to the rules of exact mathematics

That we do not know their weights, and that there are bonds between them, will not here injure the graphically determined normal places The very circumstance that even distant observations by the construction of the curve are made to influence each normal place, is an advantage It is not necessary here to suffer any loss of exactness, as by the other normal places, which, as they are to be represented as mean numbers, cannot at the same time be put in the most advantageous places and obtain the due weight As to the rest, however, what has been said p 108—110 about the necessity of putting the substitutive observations in the right place, holds good also, without any alteration, of the graphical normal places

The method of the graphical adjustment enables us to execute the drawing with absolute correctness, and it leaves us full liberty to put the normal places where we like, consequently also in the places required for absolute correctness, but in both these respects *it leaves everything to our tact and practice, and gives no formal help to it*

As to the criticism, the graphical adjustment gives no information about the mean errors of its results But, if we can state the mean error of each observation, we are able, nevertheless, to subject the graphical adjustments to a summary criticism, according to the rule

$$\sum \frac{(o-u)^2}{\lambda_2} = n - m$$

And with respect to the more special criticism on systematical deviations, the graphical method even takes a very high rank Through graphical representations of the finally remaining deviations, $o-u$, particularly if we can also lay down the mean errors on the same drawing, we get the sharpest check on the objective correctness of any adjustment.

From this reason, and owing to the proportionally slight difficulties attached to it, the graphical adjustment becomes particularly suitable where we are to lay down new empirical laws In such cases we have to work through, to check, and to reject series of hypotheses as to the functional interdependency of observations and their essential circumstances We save much labour, and illustrate our results if we work by graphical adjustment

Of course, we are not obliged to subject observations to adjustment In the preliminary stages, or as long as it is doubtful whether a greater number of essential circumstances ought not to be taken into consideration, it may even be the best thing to give the observations just as they are

But if we use the graphical form in order to illustrate such statements by the drawing of a line which connects the several observed points, then we ought to give this line the form of a continuous curve and not, according to a fashion which unfortunately is widely spread, the form of a rectilinear polygon which is broken in every observed point Discontinuity in the curve is such a marked geometrical peculiarity that it ought, even more than cusps, double-points, and asymptotes, to be reserved for those cases in which the author expressly wants to give his opinion on its occurrence in reality

XIV THE THEORY OF PROBABILITY

§ 65 We have already, in § 9, defined "*probability*" as the limit to which — the law of the large numbers taken for granted — the relative frequency of an event approaches, when the number of repetitions is increasing indefinitely, or in other words, as the limit of the ratio of the number of favourable events to the total number of trials

The theory of probabilities treats especially of such observations whose events cannot be naturally or immediately expressed in numbers But there is no compulsion in this limitation When an observation can result in different numerical values, then for each of these events we may very well speak of its probability, imagining as the opposite event all the other possible ones In this way the theory of probabilities has served as the constant foundation of the theory of observation as a whole

But, on the other hand, it is important to notice that the determination of the law of errors by symmetrical functions may also be employed in the non-numerical cases without the intervention of the notion of probability For as we can always indicate the mutually complementary opposite events as the "fortunate or "unfortunate" one, or as "Yes' and "No", we may also use the numbers 0 and 1 as such a formal indication If

then we identify 1 with the favourable "Yes"-event, 0 with the unfavourable "No", the sums of the numbers got in a series of repetitions will give the frequency of affirmative events This relation, which has been used already in some of the foregoing examples, we must here consider more explicitly

If repetitions of the same observation, which admits of only two alternatives, give the result "Yes" = 1 m times, against n times "No" = 0, then the relative frequency for the favourable event is $\frac{m}{m+n}$ But if we employ the form of the symmetrical functions for the same law of actual errors, then the sums of the powers are

$$s_0 = m+n, \quad s_1 = s_2 \quad = s_r = m \tag{121}$$

In order to determine the half-invariants by means of this, we solve the equations

$$\begin{aligned}
m &= (m+n)\mu_1 \\
m &= \quad m\ \mu_1 + (m+n)\mu_2 \\
m &= \quad m\ \mu_1 + \quad 2m\ \mu_2 + (m+n)\mu_3 \\
m &= \quad m\ \mu_1 + \quad 3m\ \mu_2 + \quad 3m\ \mu_3 + (m+n)\mu_4,
\end{aligned}$$

and find then

$$\left.\begin{aligned}
\mu_1 &= \frac{m}{m+n} \\
\mu_2 &= \frac{mn}{(m+n)^2} \\
\mu_3 &= \frac{mn(n-m)}{(m+n)^3} \\
\mu_4 &= \frac{mn(n^2-4mn+m^2)}{(m+n)^4}
\end{aligned}\right\} \tag{122}$$

Compare § 23, example 2, and § 24, example 8

All the half invariants are integral functions of the relative frequency, which is itself equal to μ_1 The relative frequency of the opposite result is $\frac{n}{m+n} = 1-\mu_1$, by interchanging m and n, none of the half-invariants of even degree are changed, and those of odd degree (from μ_3 upwards) only change their signs

In order to represent the connection between the laws of presumptive errors, we need only assume, in (122), that m and n increase indefinitely, while the probability of the event becomes $p = \frac{m}{m+n}$, and the probability of the opposite event is represented by $\frac{n}{m+n} = 1-p = q$ The half invariants are then

$$\left.\begin{aligned}
\lambda_1 &= p \\
\lambda_2 &= pq \\
\lambda_3 &= pq(q-p) \\
\lambda_4 &= pq(q^2-4pq+p^2)
\end{aligned}\right\} \tag{123}$$

Our mean values are therefore, respectively, the relative frequency and the probability itself

We must now first notice here that every half-invariant is its own fixed and simple function of the probability (the frequency). When a result of observation can be stated in the form of one single probability properly so called, we have thereby given as complete a determination of the law of errors as by the whole series of half-invariants. In such cases it is simpler to employ the theory of probability instead of the symmetrical functions and the method of the least squares.

The theory of probability thereby gets its province determined in a much more natural and suitable way than that employed in the beginning of this paragraph.

But at the same time we see that the form of the half-invariants is not only the general means which must be employed where the conditions for the use of the probability are not fulfilled, but also that, within the theory of probability itself, we shall require, particularly, the notion of the mean error.

Even where the probability can replace all the half-invariants, we shall require all the various sides of the notions which are distinctly expressed in the half-invariants. Now we have particularly to consider the probability as the definite mean value, now the point is to elicit the definite degree of uncertainty which is implied in the probability, and which is particularly emphasised in the mean error. Otherwise, we should constantly be tempted to rely on the predictions of the theory of probability to an extent far beyond what is justly due to them. Finally, we shall see immediately that the laws of error of the probabilities are far from typical, but that they have rather a type of their own, which must sometimes be especially emphasised.

All this we shall be able to do here, where we have the half-invariants in reserve as a means of representing the theory of probability.

§ 66. In particular, we can now, though only in the form of the half invariants, solve one of the principal problems of the theory of probability, and determine the law of presumptive errors for the frequency m of one of the events of a trial, which can have only two events and which is repeated N times, upon the supposition that the trial follows the law of the large numbers, and that the probability p for a single trial is known.

The equations (123) give us already the corresponding law of error for each trial, and as the total absolute frequency is the sum of the partial ones, we need only use the equations (35) to find:

$$\left.\begin{aligned}
\lambda_1(m) &= Np \\
\lambda_2(m) &= Npq = Np(1-p) \\
\lambda_3(m) &= Npq(q-p) = Np(1-p)(1-2p) \\
\lambda_4(m) &= Npq(q^2-4pq+p^2) \\
&= Np(1-p)(1-(3+\sqrt{3})p)(1-(3-\sqrt{3})p)
\end{aligned}\right\} \qquad (124)$$

The ratio of the mean frequency to the number of trials is therefore the probability itself When p is small the mean error differs little from the square root $\sqrt{Np}$ of the mean frequency, and if p is nearly $= 1$, the mean error of the opposite event is nearly equal to $\sqrt{Nq}$ When the probability, p, is nearly equal to $\frac{1}{2}$, the mean error will be about $\frac{1}{2}\sqrt{N}$

The law of error is not strictly typical, although the rational function of the r^{th} degree in $\lambda_r(m)$ vanishes for r different values of p between 0 and 1, the limits included, so that the deviation from the typical form must, on the whole, be small If, however, we consider the relative magnitude of the higher half-invariants as compared with the powers of the mean error

$$\left.\begin{aligned} \lambda_3(m)\ (\lambda_2(m))^{-\frac{3}{2}} &= \frac{q-p}{\sqrt{Npq}} \\ \text{and}\qquad \lambda_4(m)\ (\lambda_2(m))^{-2} &= \frac{q^2-4pq+p^2}{Npq} \end{aligned}\right\} \qquad (123)$$

the occurence of Npq in the denominators of the abridged fractions shows, not only that great numbers of repetitions, here as always, cause an approximation to the typical form, but also that, in contrast to this, the law of error in the cases of certainty and impossibility, when $q = 0$ and $p = 0$, becomes skew and deviates from the typical in an infinitely high degree, while at the same time the square of the mean errors becomes $= 0$ This remarkable property is still traceable in the cases in which the probability is either very small or very nearly equal to 1 In a hundred trials with the probability $= 99\frac{1}{2}$ per ct the mean error will be about $= \sqrt{\frac{1}{2}}$ Errors beyond the mean frequency $99\frac{1}{2}$ cannot exceed $\frac{1}{2}$, and are therefore less than the mean error The great diminishing errors must therefore be more frequent than in typical cases, and frequencies of 97 or 98 will not be rare in the case under consideration, though hey must be fully counter-balanced by numerous cases of 100 per ct The law of error is consequently skew in a perceptible degree In applications of adjustment to problems of probability, it is, from this reason, frequently necessary to reject extreme probabilities

XV THE FORMAL THEORY OF PROBABILITY

§ 67 The formal theory of probability teaches us how to determine probabilities that depend upon other probabilities, which are supposed to be given Of course, there are no mathematical rules specially applicable to computations that deal with probabilities, and there are many computations with probabilities which do not fall under the theory of probability, for instance, adjustments of probabilities But in view of the direct application

of probabilities, not only to games, insurances, and statistics, but to all conditions of life, it will be understood that special importance attaches to the marks which show that a computation will lead us to a probability as its result, as this implies in part or in the whole a determination of a law of errors The formal theory of probabilities rests on two theorems, one concerning the addition of probabilities, the other concerning their multiplication

I The theorem concerning *the addition of probabilities* can, as all probabilities are positive numbers, be deduced from the usual definition of addition as a putting together if a sum of probabilities is to be a probability itself, we must be allowed to look upon each of the probabilities that we are to add together as corresponding to its particular events These events must mutually exclude one another, but must at the same time have a quality in common, to which, after the addition, our whole attention must be given If the sum is to be the correct probability of events with this quality, the same quality must be found in no other event of the trial An "either—or" is, therefore, the simple grammatical mark of the addition of probabilities The event E, whose probability is $p_1 + p_2$, must occur, if *either* the result E_1, whose probability is p_1, *or* the quite different event E_2, whose probability is p_2, occurs, *and not* in any other case If we require no other resemblance between the events whose probabilities are added together, than that they belong to the same trial, their sum must be the probability 1, certainty, because then all events of the trial are favourable If p be the probability for a certain event, q the probability *against* the same, then we have $p + q = 1$, $q = 1 - p$ If n events of the same trial be equally probable, the probability of each being $= p$, then the aggregate probability of these events is $= np$

II. The theorem concerning *the multiplication of probabilities* can, as all probabilities are proper fractions, be deduced from the definition of the multiplication of fractions, according to which the product is the same proportional of the multiplicand as the multiplier is of unity Only as probabilities presuppose infinite numbers of trials, we shall commence by proving the corresponding proposition for relative frequencies

If, in $p = p_1 p_2$, p_1 is a relative frequency, it must relate to a trial T_1 which, repeated N times, has given favourable events in Np_1 cases, and if p_2, being also a relative frequency, takes the place of multiplier, then the corresponding trial T_2, if repeated Np_1 times, must have given $(Np_1)p_2$ favourable events Now in the multiplication $p = p_1 p_2$, p must be the relative frequency of the compound trials which out of the total number of N repetitions have given $Np_1 p_2$ favourable events The trials T_1 and T_2 must *both* have succeeded as conditional for the final event As the number N can be taken as large as we please, the same proposition must hold good for probabilities

The probability $p = p_1 p_2$, as the product of the probabilities p_1 and p_2, relates to the event of a compound trial, which is favourable only if both conditional trials, T_1 and T'_2, have given favourable events; first the trial T_1 must have had the event whose probability is p_1, and then the other trial T_2 must have succeeded in the event, whose probability, *on condition of success in* T_1, is p_2. However indifferent the order of the factors may be in the numerical computation it is nevertheless, if a probability is correctly to be found as the product of the probabilities of conditional events, necessary to imagine the conditional trials arranged in a definite order. To prove this very important proposition we shall suppose that both conditional trials are carried out in every case of the compound trial. Let both T_1 and T_2 have succeeded in a cases, while only T_1 has succeeded in b cases, only T_2 in c cases, and neither in d cases. Considering each of the two trials without any regard to the other, we therefore get $\frac{a+b}{a+b+c+d} = P_1$ and $\frac{a+c}{a+b+c+d} = P_2$ as the frequencies or probabilities of their favourable events. But in the multiplication for computation of the compound probability, P_1 and P_2 are applicable only as multiplicands; the correct result $p = \frac{a}{a+b+c+d}$ is found by $p = P_1 \cdot \frac{a}{a+b}$ or by $p = P_2 \cdot \frac{a}{a+c}$, according to the order in which the trials are executed, but *not* as $p = P_1 P_2$, unless $a : b = c : d$. But this proportion expresses that the frequency or probability of the trial T_2 is not affected by the event of the trial T_1. This proportionality is the mark of freedom, if we consider the multiplication of probabilities as the determination of the law of errors for a function of two observed values whose laws of errors are given.

Since impossibility is indicated by probability = 0, we see that the compound trial is impossible, if there is any of the conditional trials that cannot possibly succeed, i. e. if $p_1 = 0$ or $p_2 = 0$ in $p = p_1 p_2$. The condition of certainty (probability = 1) in a compound trial is certainty for the favourable events of all conditional trials; for as p_1 and p_2 as probabilities must be proper fractions, $p_1 p_2 = p = 1$ will be possible only when both $p_1 = 1$ and $p_2 = 1$.

Example 1. When the favourable events of all the conditional trials, n in number, have the same probability p, the compound event, which depends on the success of all these, has the probability p^n. If by every single drawing there is the probability of $\frac{1}{2}$ for "red" and $\frac{1}{2}$ for "black", the probability of 10 drawings *all* giving red will be $\frac{1}{1024}$.

Example 2. Suppose a pack of 52 cards to be so well shuffled that the probabilities of red and black may constantly be proportional to the remainder in the stock, then the probability of the 10 uppermost cards being red will be

$$= \frac{26}{52} \cdot \frac{25}{51} \cdot \frac{24}{50} \cdot \frac{23}{49} \cdot \frac{22}{48} \cdot \frac{21}{47} \cdot \frac{20}{46} \cdot \frac{19}{45} \cdot \frac{18}{44} \cdot \frac{17}{43} = \frac{\lfloor 26 \lfloor 42 \lfloor 10}{\lfloor 52 \lfloor 10 \lfloor 10} = \frac{\beta_{26}(10)}{\beta_{52}(10)} = \frac{19}{56588} = \frac{1}{2978},$$

the $\beta_n(x)$ being binomial functions.

Example 3. Compute the probability that a man whose age is a will be still alive after n years, and that he will die in one of the succeeding m years.

If we suppose that q_i is the probability that a man whose age is i will die before his next birthday, the probability that the man whose age is a will be alive at the end of n years will be

$$P_n = (1-q_a)(1-q_{a+1}) \ldots (1-q_{a+n-1})$$

The probability Q_m of his *then* dying in either one or the other of the succeeding m years will be

$$Q_m = q_{a+n} + (1-q_{a+n})\{q_{a+n+1} + (1-q_{a+n+1})[q_{a+n+2} + \ldots + (1-q_{a+n+m-2})q_{a+n+m-1}]\},$$

or

$$1 - Q_m = (1-q_{a+n})(1-q_{a+n+1}) \ldots (1-q_{a+m+n-1})$$

The required probability of death after n years, but before the elapse of $n+m$ years, is consequently $P_n Q_m = P_n - P_{n+m}$.

The most convenient form for statements of mortality is not, as we here supposed, a table of the probabilities q_i for all integral ages i, but of the absolute frequencies l_i of the men from a large (properly infinitely large) population who will reach the age of i. After this $q_i = \frac{l_i - l_{i+1}}{l_i}$, $\left(1-q_i = \frac{l_{i+1}}{l_i}\right)$ will only be a special case of the general answer

$$P_n Q_m = \frac{l_{a+n} - l_{a+n+m}}{l_a}.$$

Example 4. We imagine a game of cards arranged in such a way that each player, in a certain order, gets two cards of the well shuffled pack, and wins or loses according as the sum of the points on his two cards is eleven or not. For 5 players we use, for instance, only the cards 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 of the same colour.

What then is the probability of h players (named beforehand) getting 11 and not any of the $5-h$ others?

Secondly, what probability, r_k, is there that the k^{th} player in succession will be the first who gets 11?

Lastly, what is the probability, q, that none of the players will get 11?

It will be found perhaps that it is not quite easy to compute these probabilities directly. In such cases it is a good plan to reconnoitre the problem by first bringing out such results as present themselves quite easily and simply, without considering whether they are just those we require. In this case, for instance, we take the probabilities, p_i, that each of the first i players will get 11.

We then attack the problem more seriously, and examine if there are not any simple functions of the probabilities we have found, p_i, which may be interpreted as probabilities of the same or similar sort as those inquired after.

$$q = \frac{[illegible]}{[illegible]} = 10(p_2 - p_3) + 5(p_4 - p_1) + p_0 - p_5$$

§ 68 Repetitions of the same trial occur very frequently in problems solvable by the theory of probabilities, and should always by treated by means of a very simple and important law, the polynomial formula

Let us suppose that the various events of the single trial may be indicated by colours, and that, in the single trial, the probability of white is w, of black b, and of red r.

The probability that we shall get in $x+y+z$ trials x white, y black, and z red results, in a given order, is then

$$w^x \; b^y \; r^z$$

The number of the events of this kind that differ only in order, is the trinomial coefficient

$$r(x, y, z) = \frac{1 \cdot 2 \cdot 3 \ldots (x+y+z)}{1 \cdot 2 \ldots x \cdot 1 \cdot 2 \ldots y \cdot 1 \cdot 2 \ldots z},$$

which is the coefficient of the term $w^x \; b^y \; r^z$ in the development of $(w+b+r)^{(x+y+z)}$ And this same term

$$r(x, y, z)\, w^x \; b^y \; r^z \tag{126}$$

is the required probability of getting white x times, black y times, and red z times by $(x+y+z)$ repetitions

When the probabilities of all possible single results are known and employed, so that $w+b+r+\ldots = 1$, and when the number of repetitions is n, we must consequently imagine $(w+b+r+\ldots)^n$ developed by the polynomial theorem, and the single terms of the development will then give us the probabilities of the different possible events of the repetitions without regard to the order of succession

Example 1 If the question is of the probability of getting, in 10 trials in which there are the three possible events of white, black, and red, even numbers x, y, and z of each colour, and if the probabilities of the single events are w, b, and r, respectively, then we must retain the terms of $(w+b+r)^{10}$ which have even indices, and we thus find

$$\begin{aligned}
&w^{10}+45w^8(b^2+r^2)+210w^6(b^4+6b^2r^2+r^4)+210w^4(b^6+15b^4r^2+15b^2r^4+r^6)+\\
&+45w^2(b^8+28b^6r^2+70b^4r^4+28b^2r^6+r^8)+b^{10}+45b^8r^2+210b^6r^4+210b^4r^6+45b^2r^8+r^{10} =\\
&= \tfrac{1}{4}\{(w+b+r)^{10}+(-w+b+r)^{10}+(w-b+r)^{10}+(w+b-r)^{10}\} =\\
&= \tfrac{1}{4}\{1+(1-2w)^{10}+(1-2b)^{10}+(1-2r)^{10}\}
\end{aligned}$$

The probability, consequently, is always greater than $\frac{1}{4}$, but only a little greater, unless the probability of getting some of the events in a single trial, is very small

Example 2 Peter and Paul play at heads-or-tails (i e probability $= \frac{1}{2}$ for and against) But Peter throws with 3 coins, Paul only with 2, and the one wins who gets

the greatest number of "heads" If both get the same number of heads they throw again, as often as may be necessary What is the probability that Peter will win?

If we write for Peter's probability for and against throwing heads $p_1 = \frac{1}{2}$ and $q_1 = \frac{1}{2}$, for Paul's $p_2 = \frac{1}{2}$ and $q_2 = \frac{1}{2}$, then we should develop $(p_1+q_1)^3 (p_2+q_2)^2$, and the terms in which the index of p_1 is greater than that of p_2, are in favour of Peter, those in which the indices are equal, give a drawn game, and those in which the index of p_2 is greater than that of p_1, are in favour of Paul For the single game there is the probability

for Peter of $\frac{8}{16}$,
for a drawn game of $\frac{5}{16}$,
for Paul of $\frac{3}{16}$

As the probabilities are distributed in the same way, when they play the games over again, we need not consider the possibilities of drawn games at all, and we find $\frac{8}{11}$ as Peter's final probability

Example 8 A game which is won once out of four times, is repeated 10 times What is the probability of winning at most 2 of these?

$$\frac{551124}{1048576}.$$

§ 69 It often occurs that we inquire in a general way concerning a probability, which is a function of one or more numbers Often it is also easier to transform a special problem into such a one of a more general character, where the unknown is a whole table $p_1, p_2, p_3, \dots p_n$ of probabilities, the suffixes being the arguments of the table And then we must generally work with implicit equations, $f(p_1, \dots p_n) = 0$, particularly such as hold good for an arbitrary value of n, i e, with difference-equations Integration of finite difference-equations is indeed of so great importance in the art of solving problems of the theory of probabilities, that we can almost understand that Laplace has treated this method almost as the one to be used in all cases, in fact as the scientific quintessence of the theory of probabilities,

Since finite difference-equations like differential equations cannot as a rule be integrated by known functions, we can in an elementary treatise deal only with the simplest cases, especially such as can be solved by exponential functions, namely the linear difference-equations with constant coefficients. As to these, it is only necessary to mention here that, when

$$c_m p_{n+m} + \dots + c_0 p_n = 0 \text{ (n being arbitrary)},$$

the solution is given by

$$p_n = k_1 r_1^n + \dots + k_m r_m^n, \qquad (197)$$

where $r_1, \ldots r_m$ are the roots in the equation

$$c_m r^m + \ldots + c_0 = 0,$$

while $k_1, \ldots k_m$ are integration-constants whenever the corresponding roots occur singly, but rational integral functions with arbitrary constants, and of the degree $i-1$, if the corresponding root occurs i times

I shall mention one other means, however, not only because it can really lead to the integration of many of the difference-equations which the theory of probabilities leads to, particularly those in which the exponential functions occur in connection with binomial functions and factorials, but also because it has played an important part in the conception of this book

The late Professor L Oppermann, in April 1871, communicated to me a method of transformation, which I shall here state with an unessential alteration

A finite or infinite series of numbers

$$u_0, u_1, \ldots u_n$$

can univocally be expressed by another

$$\left.\begin{aligned}
w_0 &= u_0 + u_1 + u_2 + u_3 + u_4 + \ldots \\
w_1 &= \quad - u_1 - 2u_2 - 3u_3 - 4u_4 - \ldots \\
w_2 &= \qquad\quad u_2 + 3u_3 + 6u_4 + \ldots \\
w_3 &= \qquad\qquad - u_3 - 4u_4 - \ldots \\
w_4 &= \qquad\qquad\qquad u_4 + \ldots \\
w_r &= (-1)^r \Sigma \beta_p(r) u_p ,
\end{aligned}\right\} \qquad (128)$$

where the sum Σ may be taken from $-\infty$ to $+\infty$, provided that $u_p = 0$ when $p > n$. In order, vice versâ, to compute the u's by means of the ws, we have equations of just the same form

$$\left.\begin{aligned}
u_0 &= w_0 + w_1 + w_2 + w_3 + w_4 + \ldots \\
u_1 &= \quad - w_1 - 2w_2 - 3w_3 - 4w_4 - \ldots \\
u_2 &= \qquad\quad w_2 + 3w_3 + 6w_4 + \ldots \\
u_3 &= \qquad\qquad - w_3 - 4w_4 - \ldots \\
u_4 &= \qquad\qquad\qquad w_4 + \ldots \\
u_p &= (-1)^p \Sigma \beta_r(p) w_r
\end{aligned}\right\} \qquad (129)$$

Here as in (17) and (18), the general dependency between the u_i and w_j can be expressed in a single equation, be means of an independent variable z. From (129) we get identically

$$u_0 + u_1 e^z + u_2 e^{2z} + \ldots = w_0 + (1-e^z) w_1 + (1-e^z)^2 w_2 + \ldots$$

If we here put $1 - e^z = e^y$, then $1 - e^y = e^z$ will reduce (128) to an equation of the same form

If u_i is the frequency or probability of i taken as an observed value, then also

$$u_0 + u_1 e^t + u_2 e^{2t} + \dots = s_0 + \frac{s_1 t}{1} + \frac{s_2 t^2}{1 \cdot 2} + \dots =$$
$$= s_0 e^{\frac{\mu_1 t}{1} + \frac{\mu_2 t^2}{1 \cdot 2} + \dots} = w_0 + (1-e^t) w_1 + (1-e^t)^2 w_2 + \dots$$

illustrate the relations of the values in Oppermann's transformation to the half-invariants and sums of powers. In particular we have

$$\mu_1 = -\frac{w_1}{w_0}$$
$$\mu_2 = 2\frac{w_2}{w_0} - \frac{w_1(w_0+w_1)}{w_0^2}$$
$$\mu_3 = -6\frac{w_3}{w_0} + 6\frac{w_2(w_0+w_1)}{w_0^2} - \frac{w_1(w_0+w_1)(w_0+2w_1)}{w_0^3}.$$

If now $u_0, u_1, \dots u_n$ are a series of probabilities or other quantities which depend on their suffix according to a fixed law, and if we know this law only through a difference-equation, then Oppermann's transformation of course leads only to a difference-equation for $w_0, w_1, \dots w_n$ as function of their suffix. But it turns out that, in problems of probabilities, this equation pretty often is easier to deal with than the original one (for instance the more difficult ones in Laplace's collection of problems). If we can look upon a probability u_i as the functional law of errors for i as the observed value, then w expresses the same law of errors by symmetrical functions, and frequently we want nothing more. If we have to reverse the process to find u_i itself, the series are pretty simple if w is simple; but they are often less favourable for numerical computation, as they frequently give the unknown as a difference between much larger quantities. There exists a means of remedying this, but it would carry us too far to enter into a closer examination of the question here.

Example 1. I throw a die, and go on throwing till I either win by getting "one" twice, or lose by throwing "two" or "three". If the game is to be over at latest by the nth throw, what is my probability of winning? If the number of throws is unlimited, what is the probability of another "one" appearing before any "two" or "three"?

Four results are to be distinguished from one another. At any throw, say the ith, the game can in general be won, lost, half won (by only one "one"), or drawn. Let the probability of the ith throw resulting in a win be p_i, of the same resulting in a loss be q_i, in half win s_i, and in a drawn game be r_i, then $p_1 = 0$, $q_1 = \frac{1}{3}$, $s_1 = \frac{1}{6}$, and $r_1 = \frac{1}{2}$. Thus the probability of a second throw is $\frac{2}{3}$, and, generally, the probability of an ith throw $s_{i-1} + r_{i-1}$. It is easy to express p_i, q_i, r_i, and s_i in terms of r_{i-1} and s_{i-1}, and also

p_{t-1}, q_{t-1}, r_{t-1}, and s_{t-1} in terms of r_{t-1} and s_{t-1}, etc. By elimination then the difference equations can be found.

When we replace p or q or s or r by x the difference equation can be written in the common form

$$x_t - x_{t-1} + \tfrac{1}{4}x_{t-2} = 0,$$

which is integrated as

$$x_t = (a + bt)2^{-t};$$

for r we have the simpler form

$$r_t = \tfrac{1}{2}r_{t-1}.$$

When, by the probabilities of the first throws, we have determined the constants, we get

$$p_t = \frac{t-1}{9}2^{-t},$$

$$q_t = \frac{2t+4}{9}2^{-t},$$

$$s_t = \frac{t}{3}2^{-t},$$

and

$$r_t = 2^{-t}.$$

We then have the formulae $P_n = p_1 + \ldots + p_n$ and $Q_n = q_1 + \ldots + q_n$, for the probabilities of making the winning or losing throw, and we get

$$\frac{P_n}{P_n + Q_n} = \frac{1}{3}\,\frac{(2^n-1)-n}{3(2^n-1)-n} \quad \text{and} \quad \frac{P\infty}{P\infty + Q\infty} = \frac{1}{9}.$$

Example 2. In a game the probability of winning is ϖ. The same game is repeated a great many, n, times. If it then happens at least once in this series that in succession e games are won, you get a prize. What is the probability p_n of this? In a game of dice, where $\varpi = \frac{1}{6}$, what is the probability of getting a series of 5 "sixes" in 10000 throws?

It will be simplest to find the probability, $q_r = 1 - p_r$, that the prize will not be got in the first r repetitions. The difference-equation for this is

$$q_{r+m+1} - q_{r+m} + (1-\varpi)\varpi^m q_r = 0 \qquad \text{(a)}$$

or

$$\varpi^r c_0 = q_{r+m} - (1-\varpi)\{q_{r+m-1} + \varpi q_{r+m-2} + \ldots + \varpi^{m-2} q_{r+1} + \varpi^{m-1} q_r\} = 0, \quad \text{(b)}$$

where (b) is the first integral of (a). (As well as (a) we can directly demonstrate (b). How?). Hence

$$q_r = c_1\rho_1^r + \ldots + c_m\rho_m^r,$$

where $c_1, \ldots c_m$ are constants, which as well as $c_0 = 0$ must be determined by means of

$q_0 = q_1 = \cdots = q_{m-1} = 1$, $q_m = 1 - \varpi^m$, and ρ_1 to ρ_m are the roots of an irreducible equation of the m^{th} degree, which is got from

$$\rho^{m+1} - \rho^m = \varpi^{m+1} - \varpi^m \tag{c}$$

by dividing out $\rho - \varpi$. The largest of these roots (for small ϖ's or large m's) will be only a little less than 1, a small negative root occurs when m is even, the others are always imaginary, and they are also small.

In the actual computation it is highly desirable to avoid the complete solution of (c). This can be done, and this problem will illustrate a most important artifice. We may use the difference-equation to compute a single value of the unknown function by means of those which are known to us from the conditions of the problem, and then successive values of the unknown function by means of those already obtained; here, for instance, (b) enables us to get q_{m+1} in terms of $q_1, \ldots q_m$. Then we get q_{m+2}, either by again applying (b) to $q_2, \ldots q_{m+1}$, or by applying (a) to q_1 and q_{m+1} (or best in both ways for the sake of the check), etc.

It is evident that the table of the numerical values of the function which we can form in this way, cannot easily become of any great extent or give us exact information as to the form of the function. But we are able to interpolate, and, when the general form of the function is known (as here), we may be justified in using extrapolation also. In our example we need only continue the computations above described until the term in $q_r = c_1\rho_1^r + \cdots$, corresponding to the greatest root ρ_1, dominates the others to such a degree that the first difference of $\operatorname{Log} q_r$ becomes constant, and the computation of q_r for higher indices can then be made as by a simple geometrical progression. In the numerical case $q_r = 1.004078 \times (0.9998928)^r$, $1 - q_{1000} = 0.6577$.

Example 9. A bag contains n balls, a white and $a - n$ black ones. A ball is drawn out of the bag and a black ball then placed in it, and this process is repeated y times. After the y^{th} operation the white and black balls in the bag are counted. Find the probability $u_x(y)$ that the numbers of white balls will then be x and the black ones $n - x$.

We have

$$u_x(y) = \frac{n-x}{n} u_x(y-1) + \frac{x+1}{n} u_{x+1}(y-1)$$

and

$$u_x(0) = 0, \text{ except } u_a(0) = 1.$$

By Oppermann's transformation we find

$$w_z(y) = (-1)^z \Sigma \beta_x(z) \cdot u_x(y),$$

Σ taken from $x = -\infty$ to $x = +\infty$, or

$$w_z(y) = (-1)^z \Sigma \frac{n-x}{n} \beta_x(z) \cdot u_x(y-1) + (-1)^z \Sigma \frac{x+1-z}{n} \beta_{x+1}(z) \cdot u_{x+1}(y-1)$$

The limits of z under Σ being infinite, $z+1$ can be replaced by z, consequently

$$w_z(y) = \frac{n-z}{n} w_z(y-1)$$

This difference-equation, in which y is the variable, may easily be integrated As we have, further

$$w_z(0) = (-1)^z \beta_a(z),$$

we get

$$w_z(y) = (-1)^z \ \beta_a(z) \left(\frac{n-z}{n}\right)^y$$

By Oppermann's inverse transformation we find now

$$u_x(y) = (-1)^x \Sigma \beta_x(z) \ (-1)^z \ \beta_a(z) \cdot \left(\frac{n-z}{n}\right)^y,$$

Σ taken from $z = -\infty$ to $z = +\infty$ This expression

$$u_x(y) = \beta_a(x) \Sigma (-1)^{x+z} \ \beta_{a-x}(z-x) \ \left(\frac{n-z}{n}\right)^y$$

has the above mentioned practical short-comings, which are sensible particularly if n, $a-x$, or y are large numbers, In these cases an artifice like that used by Laplace (problem 17) becomes necessary But our exact solution has a simple interpretation The sum that multiplies $\beta_a(x)$ in $u_x(y)$, is the $(a-x)^{\text{th}}$ difference of the function $\left(\frac{n-z}{n}\right)^y$, and is found by a table of the values $\left(\frac{n-a}{n}\right)^y$, $\left(\frac{n-a+1}{n}\right)^y$, … $\left(\frac{n-x-1}{n}\right)^y$, $\left(\frac{n-x}{n}\right)^y$, as the final difference formed by all these consecutive values We learn from this interpretation that it is possible, if not easy, to solve this problem without the integration of any difference-equation, in a way analogous to that used in § 67, example 4.

If we make use of $w_z(y)$ to give us the half-invariants μ_1, μ_2 for the same law of errors as is expressed by $u_x(y)$, then we find for the mean value of x after y drawings

$$\lambda_1(y) = a\left(\frac{n-1}{n}\right)^y$$

and for the square of the mean error

$$\lambda_2(y) = a\left(\left(\frac{n-1}{n}\right)^y - \left(\frac{n-2}{n}\right)^y\right) + a^2\left(\left(\frac{n-2}{n}\right)^y - \left(\frac{n-1}{n}\right)^{2y}\right)$$

XVI. THE DETERMINATION OF PROBABILITIES A PRIORI AND A POSTERIORI

§ 70 The computations of probabilities with which we have been dealing in the foregoing chapters have this point in common that we always assume one or several probabilities to be given, and then deduce from them the required ones If now we ask, how

we obtain those "given" probabilities, it is evident that other means are necessary than those which we have hitherto been able to mention, and provisionally it must be clear that both theory and experience must cooperate in these original determinations of probabilities Without experience it is impossible to insure agreement with reality, and without theory in these as well as in other determinations we cannot get any firmness or exactness In determining probabilities, however, there is special reason to distinguish between two methods, one of which, the *a priori* method seems at first sight to be purely theoretical, while the other, the *a posteriori* method, is as purely empirical

§ 71 The *a priori* determination of probabilities is based on estimate of equality, inequality, or ratio of the probabilities of the several events, and in this process we always assume the operative causes, or at any rate their mode of operation, to be more or less known

On the one hand we have the typical cases in which we know nothing else with respect to the events but that each of them is possible, and in the absence of any reason for preferring any one of them to any other, we estimate them to be equally probable — though certainly with the utmost uncertainty For instance What is the probability of seeing, in the course of time, the back of the moon? Shall we say $\frac{1}{2}$ or $\frac{2}{3}$?

On the other hand we have the cases — equally typical, but far more important — in which, by virtue of a good theory, we know so much of the causes or combinations of causes at work that, for each of those which will produce one event, we can point out another (or n others) which will produce the opposite event, and which according to the theory must occur as frequently In this case we must estimate the probability of the result at $\frac{1}{2}$ and $\frac{1}{n+1}$ respectively, and if the conditions stated be strictly fulfilled, such a determination of probability will be exact

But even if such a theory is not absolutely unimpeachable, we can often in this way obtain probabilities, which are so nearly exact and have such infinitely small mean errors, that we may very well make use of them, and compute from them values which may be used as our theoretically given probabilities We are not more strict in other kinds of computations In astronomical adjustment, for instance it is almost an established practice to consider all times of observation as theoretically given Their real errors, however, will often give occasion to sensible bonds between the observed co-ordinates; but the fact is that it would require great labour to avoid the drawback

Such an a priori determination of probabilities is particularly applicable in games For it is essential to the idea of a game that the rules must be laid down in such a way that, on the one hand they exclude all computation beforehand of the result in a particular case, while on the other hand they make a pretty exact computation of the probabilities possible The procedure employed in a game, e g. throwing of dice or shuffling

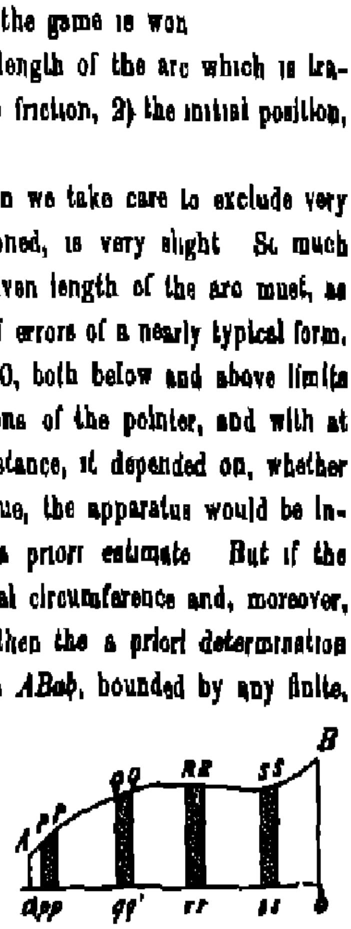
of cards, ought therefore to exclude all circumstances that might permit the players to set causes in train, which could bring about or further a certain event (corriger la fortune). But also those circumstances ought to be eliminated, which not only by their incalculability make a judgment of the probabilities very insecure, but, above all, make it depend on the theoretical insight of the parties Otherwise the game will cease to be a fair game and will become a struggle The so-called stock-jobbing is rather a war than a game

When the estimate of the probabilities depends essentially upon personal knowledge, we speak of a *subjective probability* This too plays a great part, especially in daily life The fear which ignorant people have of all that is new and unknown, proves that they understand that there is a great uncertainty in the estimate, and that it is greater for those who know but a little, than for those who know more and are therefore better able to judge

Roulette may be taken as an example of the *objective probability* which arises in a well arranged game A pointer turns on an almost frictionless pivot and points to the scale of a circle whose center is in the pivot The pointer is made to revolve quickly, and the result of the game depends on where it stops If the pointer stops opposite a space — suppose a red one — previously selected as favourable, the game is won

There we have as essential circumstances 1) the length of the arc which is traversed, this being determined by the initial velocity and the friction, 2) the initial position, and 3) the manner in which the circle is divided

The length of the arc is unknown, especially when we take care to exclude very small velocities, and when the friction, as already mentioned, is very slight So much only may be regarded as given, that the frequency of a given length of the arc must, as function of this length, be expressed by a functional law of errors of a nearly typical form. For the frequency must go down, asymptotically, as far as 0, both below and above limits of the arc which will be separated by many full revolutions of the pointer, and with at least one maximum between these limits If now, for instance, it depended on, whether the arc traversed was greater or smaller than a certain value, the apparatus would be inexpedient, it would not allow any tolerably trustworthy a priori estimate But if the winning space (or spaces) is small in proportion to the total circumference and, moreover, repeated regularly for each of the numerous revolutions, then the a priori determination of the probabilities will be even very exact For an area *ABab*, bounded by any finite, continuous curve whatever (in the present case the curve of errors of the different possible events), by the axis of abscissae, and two ordinates, can always as a first approximation be expressed as the sum of numerous equidistant small areas *pP*, *qQ* . with a constant base, multiplied by the



interval $pq = qr =$ and divided by the base $pp' = qq' =$ And if we speak of the total area of a *curve of errors*, then the series of which the first term is this approximation, is even very convergent, in such a degree as $\theta(x) = 1 + x + x^4 + x^9 + x^{16} +$ for small x, and the said approximation is sufficient for all practical purposes

That the initial position of the roulette is unknown, does not essentially change the result of the foregoing, viz that the probability of winning is $\frac{pp'}{pq}$ This uncertainty can only cause an improvement of the accuracy of this approximation. If we may assume that the pointer will as probably start from any point in the circle as from any other, this determination $\frac{pp'}{pq}$ will even be exact, without any regard to the special kind of the unknown function of frequency

The ratio of the winning space on the circle pp' to the whole circumference pq, the third essential circumstance, cannot be determined wholly a priori, but demands a measurement or a counting whose mean error it is essential to know

The a priori determination of probability can thus, according to circumstances, give results of the most different values, from the very poorest through gradual transition up to such exact probabilities as agree with the suppositions in § 65 seqq, and permit the probability to replace the whole law of errors for our predictions. But what the a priori method cannot give, is a quantitative statement of the uncertainty which affects the numerical value of the probability itself. Only when it is evident, as in the example of the roulette, that this uncertainty is infinitely small, can we make use of a priori probabilities in computations that are to be relied on If in the work and struggles of our life, we cannot entirely avoid building on altogeth r uncertain and subjective a priori estimates, great caution is necessary, and in order not to overdo this caution for want of a proper measure, we must try, by tact or experience, without any real method, to get an estimate of the uncertainty.

Even by the best a priori determinations of probability caution is not superfluous; the dice may be false, the pivot of the roulette may be worn out or bent, and so on

§ 72 By the *a posteriori determination of probability* we build on the law of the large numbers, inferring from a law of actual errors in the form of frequency to the law of presumptive errors in that of the probability We repeat the trial or the observation, and count the numbers m for the favourable and n for the unfavourable events

Owing to the signification of a probability as mean value, the single values being 0 for every unfavourable event, 1 for every favourable event, the probability p for the favourable event must be transferred unchanged from the law of actual errors to that of presumptive errors; consequently

$$p = \frac{m}{m+n} \tag{180}$$

Since, according to the same consideration the square of the mean deviation for a single trial is $\frac{s_2 s_0 - s_1^2}{s_0^2} = \frac{mn}{(m+n)^2}$, and the number s_0 of the repetitions is $= m+n$, the square of the mean errors must, according to (47), be

$$\lambda_2 = \frac{mn}{(m+n)(m+n-1)}, \tag{131}$$

which is, therefore, the square of the mean error for a single trial whether this is one of those which we have made, or is a repetition which we are still to make and for which we are to compute the uncertainty

If we then ask for the mean error of the probability $p = \frac{m}{m+n}$, got from the $m+n$ repetitions, we have

$$\lambda_1(p) = \frac{mn}{(m+n)^2(m+n-1)} = \frac{p(1-p)}{m+n-1} \tag{132}$$

as the square of this mean error

The identity

$$\frac{mn}{(m+n)^2} + \frac{mn}{(m+n)^2(m+n-1)} = \frac{mn}{(m+n)(m+n-1)}$$

or

$$pq + \lambda_2(p) = \lambda_2 \tag{133}$$

shows that the mean error at a single trial, when the probability p is determined a posteriori by $m+n$ repetitions, can be computed by (34), as originating in two mutually free sources of errors, one of which is the normal uncertainty belonging to the probability, for which $\lambda_2 = pq$ (129), while the other is the inaccuracy of the a posteriori determination, for which $\lambda_2(p)$ is the square of the mean error

The a posteriori determination therefore never gives an exact result, but only an approximation to the probability Only when the number of repetitions we employ is so large that their reduction by a unit may be regarded as insignificant, we can immediately employ the probabilities found by means of them as complete expressions for the law of errors But even by the very smallest number of repetitions of the trial, we not only obtain some knowledge of the probability, but also a determination of the mean error, which may be useful in predictions, and may serve as a measure of the caution that is necessary It must be admitted that it is not such a simple thing to employ these mean errors as those in the ideal theory of probability, but it is not at all difficult

As above mentioned, the a posteriori determination of probability seems to be purely empiric; theory, however, takes part in it, but is concealed in the demand, that all the trials we make use of must be repetitions, in the same way as the future trials whose results and uncertainty are predicted by the a posteriori probabilities Transgressions of this rule, which reveal themselves by unsuccessful predictions, are by no means rare, and compel statistics and the other sciences which work with probabilities, to many alterations

of their theories and hypotheses, and to the division of the materials obtained by trial into more and more homogeneous subdivisions

Example A die is inaccurate and suspected of being false On trial, however, we have on throwing it 126 times got "six" exactly 21 times, and so far, all is right. The probability of "six" is found, consequently, to be $p = \frac{21}{126} = \frac{1}{6}$; the square of the mean error is $\lambda_2(p) = \frac{1}{6}\ \frac{5}{6}\ \frac{1}{125} = \frac{1}{900}$; the limits indicated by the mean errors are consequently $\frac{1}{6} \pm \frac{1}{30}$, or $\frac{2}{15}$ and $\frac{1}{5}$.

If now we seek the probability that we shall not get "six" in 6 throws, the probability is still as by an accurate die $(1-p)^6 = \frac{15625}{46656} = \frac{1}{3} + \ldots$, but what is now the mean error? Ideally, its square should be $(1-p)^6(1-(1-p)^6) = \frac{2}{9} + \ldots$ But if p can have a small error dp, the consequent error in $(1-p)^6$ will be $-6(1-p)^5\,dp$; if then the square of the mean error of p is $= \frac{1}{900} = p(1-p)\frac{1}{s_2-1}$, the total square of the mean error of the probability of not getting "six" in 6 throws will be

$$\lambda_2 = (1-p)^6(1-(1-p)^6) + 36(1-p)^{10}\ p(1-p)\frac{1}{s_2-1}$$
$$= \frac{2}{9} + \quad + \frac{2}{311} + \quad = \frac{8}{35} +$$

In every single game of this sort the mean error is therefore only slightly larger than with an accurate die, but its actual value is so large (nearly $\frac{1}{2}$) as to call for so much caution on the part both of the player and of his opponent, that there is not much chance of their laying a wager This may be remedied by stipulating for a large number of repetitions of the game Let us examine the conditions if we are to play this game of making 6 throws without "six" 72 times With the above approximate fractions there will be expectation of winning $72\ \frac{1}{3} = 24$ games. In the computation of the square of the mean error of this result, the first term in the above λ_2 must be multiplied by 72, but the second by 72^2, hence

$$\lambda_2 = \frac{2}{9}\ 72 + \frac{2}{311}\ 5184$$
$$= 16 + 33 = 49$$

The mean error will be about 7, while it would only have been 4, if the die had been quite trustworthy

§ 73 We have mentioned already, in § 66, the skewness of the laws of errors which is peculiar to all probability It does not disappear, of course, in passing from the law of actual errors to that of presumptive errors, and in the a posteriori determination of probability it produces what we may call the *paradox of unanimity* If all the repetitions we have made agree in giving the same event, the probability deduced from this, a posteriori, must not only be 1 or 0, but the square of the mean error $\lambda_2(p)$ of these determi-

nations (as well as the higher half-invariants) becomes $= 0$ Must we infer then, respectively, to certainty or to impossibility, only *because a smaller or greater number of repetitions* mutually agree? must we consider a unanimous agreement as a proof of the absolute *correctness* of that which is thus agreed upon? Of course not; nor can this inference be maintained, if we look more closely at the law of errors $\mu_1 = 0$, $\mu_2 = 0$, $\ldots$ $\mu_n = 0$ Such a law of errors, to be sure, *may* signify certainty, but not when, as here, the ratio $\mu_3 : \mu_2^{\frac{3}{2}} = \infty$ A law of errors which is skew in an infinitely high degree, must indicate something peculiar, even though the mean error be ever so small Add to this that it is not a *strict consequence in practical calculations* that, because the *square of a number*, here that of the mean error, is $= 0$, the number itself must be $= 0$, but only that it must be so small that it may be treated as a differential, which otherwise is indeterminate The paradox being thus explained, it follows that no objections against the use of a posteriori probabilities in general can be based on it. But it must warn us to be cautious in computations with such probabilities as observed values, where the computation, as the method of the least squares, presupposes typical laws of errors For this reason, we must *for such computations reject all unanimously or nearly unanimously* determined probabilities as unsuitable material of observation Another thing is that we must also reject the hypothesis or *theory* of the computation, if it does not explain the unanimity As an example we may take an examination of the probability of marriage at different ages. The a posteriori statistics before the c 20th year and after the c 60th must not be used in the computation of the sought constants of the formula, but the formula can be employed only when it has the quality of a functional law of errors so that it approaches asymptotically towards 0, both for low and high ages

The paradox of unanimity has played rather a considerable part in the history of *the theory of probabilities It has even been thought that we ought to compute a posteriori* probabilities by another formula

$$p = \frac{m+1}{m+n+2} \text{ (Bayes's Rule)} \tag{134}$$

and not, as above, by the formula of the mean number

$$p = \frac{m}{m+n}$$

The proofs some authors have tried to give of Bayes's rule are open to serious objections In the "*Tidsskrift for Mathematik*" *(Copenhagen, 1879), Mr Bing has given a crushing* criticism of these proofs and their traditional basis, to which I shall refer those of my readers who take an interest in the attempts that have been made to deduce the theory of probabilities mathematically from certain definitions

Bayes's rule has not been employed in practice to any greater extent, particularly not in statistics, though this science works entirely with a posteriori probability. But as it makes the paradox of unanimity disappear in a convenient way, and as, after all, we can neither prove nor disprove the exact validity of a formula for the determination of an a posteriori probability, any more than we can do so for any transition whatever from the law of actual errors to that of presumptive errors, the rule certainly deserves to be tested by its consequences in practice before we give it up altogether The result of such a test will be that the hypothesis that Bayes's rule will give the true probability, can never deviate more than at most the amount of the mean error from the result of the series of repetition, viz that m events out of $m+n$ have proved favourable In order to demonstrate this proposition we shall consider a somewhat more general problem

If we assume that trials have been previously made which have given μ favourable, ν unfavourable events, and that we have now in continuing the trials found m favourable and n unfavourable events, then the probability, being looked upon as the mean value, is determined by

$$p = \frac{m+\mu}{m+n+\mu+\nu}, \tag{135}$$

of which Bayes's formula is the special case corresponding to $\mu = \nu = 1$. Bayes's rule would therefore agree with the general rule, if we knew before the a posteriori determination so much of the probability of both cases, as a report of one earlier favourable event and one unfavourable event

In the more general case the square of the mean error at the single trial is now

$$\lambda_2 = \frac{(m+\mu)(n+\nu)}{(m+n+\mu+\nu)(m+n+\mu+\nu-1)},$$

and for the $m+n$ trials is

$$\lambda_2(m+n) = (m+n)\lambda_2$$

If we now compare with this the square of the deviation between the new observation and its computed value, that is, between m and $(m+n)p$, we find

$$\frac{(m-(m+n)p)^2}{\lambda_2(m+n)} = \frac{(\mu n-\nu m)^2}{(m+\mu)(n+\nu)(m+n)}\cdot\frac{m+n+\mu+\nu-1}{m+n+\mu+\nu}$$

$$= (\mu+\nu)\left(\frac{\mu}{\mu+\nu}-\frac{m}{m+n}\right)\left(\frac{\mu}{m+\mu}-\frac{\nu}{n+\nu}\right)\frac{m+n+\mu+\nu-1}{m+n+\mu+\nu} \tag{136}$$

It appears at once from the latter formula that the greatest imaginable value of the ratio is the greatest of the two numbers μ and ν In Bayes's rule $\mu = \nu = 1$ Here, therefore, 1 is the absolute maximum of the ratio of the square of deviation to that of the mean error With respect to Bayes's rule the postulated proposition is hereby demonstrated But at the same time it will be seen that we can replace Bayes's rule by a better one, if there is

only an a priori determination, however uncertain, of the probability we are seeking If we take the a priori probabilities ϖ for, and $(1-\varpi)$ against, instead of μ and ν, so that

$$p = \frac{m+\varpi}{m+n+1}, \tag{137}$$

then we are certain to avoid the paradox of unanimity where it might do harm, without deviating so much as the mean error from the observation in the a posteriori determination

Neither Bayes's rule nor this latter one can be of any great use, but we can always *employ them, when the found probabilities can be looked upon as definitive results* On the other hand, the formula of the mean value *may* be used in all cases, if we interpret the *paradox of unanimity correctly* *Where the found probabilities are to be subjected to* adjustment, the latter formula, as I have said, *must* be employed, nor can the other rules be of any help in the cases where observed probabilities have to be rejected on account of the skewness of the law of errors

XVII MATHEMATICAL EXPECTATION AND ITS MEAN ERROR

§ 74 Whether the theory of probability is employed in games, in insurances, or elsewhere, in all cases nearly in which we can speak of a favourable event, the prediction of the practical result is won through a computation of the mathematical expectation The gain which a favourable event entails, has a value, and the chance of winning it must as a rule be bought by a stake The question is How are we to compare the value of the latter with that of which the game gives us expectation? Imagine the game to be repeated, and the number of repetitions N to become indefinitely large, then it is clear, according to the definition of probability, that the sum of the prizes won, if each of them is V, must be pNV, when p indicates the probability The gain to be expected from every single game is consequently pV, and this product *of the probability* and the value of the prize is what we call mathematical expectation

The adjective *"mathematical" warns us not to consider pV as the real value which* the possible gain has for a single player This value, certainly, depends, not only objectively *on the quantity of good things* which *form* the prize, but also on purely subjective circumstances, among others on how much the player previously possesses and requires of the same sort of good things. An attempt which has been made to determine by means of what is called the "moral expectation", whether a game is advantageous or not, must certainly be regarded as a failure For it takes into account the probable change in the logarithm of

the player's property, but it does not take into consideration his requirements and other subordinate circumstances We shall not here try to solve this difficulty.

It is evident, with respect to the mathematical expectation, that if we play several unbound games at the same time the total mathematical expectation is equal to the sum of that of the several games The same is the case, if we play a game in which each event entitles the player to a special (positive or negative) prize In this latter case we speak of the total mathematical expectation as made up of partial ones

Example 1. We play with a die in such a way that a throw of 1 or 2 or 3 wins nothing, a throw of 4 or 5 wins 2 s., and one of 6 wins 8 s The total mathematical expectation is then $\frac{1}{2} \times 0 + \frac{1}{3} \times 2 + \frac{1}{6} \times 8 = 2$ s. A stake of 2 s will consequently correspond to an even game. We might also deduce the 2 s throughout, so that a throw of 1, or 2, or 3, causes a loss of 2 s and a throw of 6 a gain of 6 s, the total mathematical expectation then becomes $= 0$

Example 2 In computations of the various kinds of life-insurances the basis is 1) the table of the number of persons $l(a)$ living at a given age a The probability of such a person living x years is $= \frac{l(a+x)}{l(a)}$, of his dying within x years $= \frac{l(a)-l(a+x)}{l(a)}$, of his dying at the exact age of $a+x$ years $= -\frac{dl(a+x)}{l(a)\,dx}dx$, and from these all other necessary probabilities may be found, 2) the rate of interest ρ, which serves for the valuation of future payments of capital, $(1+\rho)^{-x}V$, or annuities certain $(1-(1+\rho)^{-x})\frac{v}{\rho}$

The value of an endowment of capital, V, payable in x years, if the person who is now a years old is then alive, is thus equal to the mathematical expectation

$$V\frac{l(a+x)}{l(a)}(1+\rho)^{-x} = \frac{l(a+x)(1+\rho)^{-(a+x)}}{l(a)(1+\rho)^{-a}}V = \frac{D(a+x)}{D(a)}V, \tag{138}$$

which, as we see, is most easily computed by means of a table of the function

$$D(z) = l(z)(1+\rho)^{-z}$$

Such a table is of great use for other purposes also

The value of an annuity, v, due at the end of every year through which a person now a years old shall live, can be computed as a sum of such payments, or by the formula

$$v\sum_{x=1}^{x=\infty}\frac{l(a+x)}{l(a)}(1+\rho)^{-x} = \frac{v}{D(a)}\sum_{x=1}^{x=\infty}D(a+x), \tag{139}$$

where $l(\infty)=0$ and $D(\infty)=0$

But it deserves to be mentioned that this same mathematical expectation is most safely looked upon as a total mathematical expectation in a game whose events are the various possible years of death, the probability of death in the first year being $\frac{l(a)-l(a+1)}{l(a)}$, in

the second $\frac{l(a+1)-l(a+2)}{l(a)}$, and so on, while the corresponding values are annuities certain of v for varying duration. In this way we find for the value of the life-annuity the expression

$$\frac{v}{\rho\, l(a)} \sum_{x=1}^{x=\infty} \left(l(a+x) - l(a+x+1)\right)\left(1-(1+\rho)^{-x}\right) \tag{140}$$

Since the sum $\sum_{1}^{\infty}(l(a+x)-l(a+x+1)) = l(a)$, we find by solution of the last parenthesis that the expression may be written

$$\frac{v}{\rho} - \frac{v}{\rho}\,\frac{1}{l(a)} \sum_{x=1}^{x=\infty} (l(a+x) - l(a+x+1))(1+\rho)^{-x},$$

and this shows that the value of the life-annuity is the difference between the capital sum of which the yearly interest is v and the value of a life-insurance of $\frac{v}{\rho}$ payable at the beginning *of the year of death*.

In life-insurance computations integrals are often employed with great advantage, instead of the sums we have used here; periodical payments (yearly, half-yearly, or quarterly) being reduced to continuous payments, and vice versâ.

§ 75. That mathematical expectation is not a solid value, but an uncertain claim, is expressed in the law of errors for the mathematical expectation, and particularly in its mean error; for owing to the frequent repetitions and combinations in games and insurances, it does not matter much that the isolated laws of errors, here as for the probabilities, are often skew. If the value V is given free of error, the square of the mean error of the mathematical expectation, $H = pV$, is, according to the general rule, to be computed by

$$\lambda_2(H) = p(1-p)V^2 \tag{141}$$

If there are N repetitions of the same game we get

$$H' = pNV$$

and

$$\lambda_2(H') = p(1-p)NV^2, \tag{142}$$

and for the total expectation of mutually free games, $H'' = \Sigma p_i N_i V_i$, we have

$$\lambda_2(H'') = \Sigma p_i(1-p_i) N_i V_i^2 \tag{143}$$

By free games we may pretty safely understand such as are not settled by the various events of the same trial or game. (As to these, see § 76.)

The mean error is excellently adapted for computing whether we ought to enter upon a proposed game, or how highly we are to value uncertain claims or outstanding balance of accounts. Such things of course are regulated by the boldness or caution of the person concerned, *but even the most cautious man may under fairly typical circum-*

stances be contented with diminishing the value of his mathematical expectation by 4 times the amount of the mean error, and it would be sheer foolhardiness, if a passionate player would venture a stake which exceeded the mathematical expectation by the quadruple of its mean error. On the other hand, a simple subtraction or addition of the mean error cannot be counted a very strong proof of caution or boldness respectively.

Example 1 A game is arranged in such a way that the probability of winning from the person who keeps the bank is $\frac{1}{10}$, the prize is 8 $ In n games the mathematical expectation with mean error is then $(0{\cdot}8\,n \pm 2{\cdot}4\sqrt{n})$ $ If the banker has no property, but may expect 144 games to be played before the prizes are to be paid, he cannot without imprudence estimate his negative mathematical hope, his fear, lower than $0{\cdot}8 \times 144 + 2{\cdot}4 \times 12 = 144$ $ He must consequently fix the stake for each game at about one dollar, and will thus stand a chance of seeing the bank broken about once in six times If, however, he has got so much capital or credit, as also so many customers, that he can play about 2304 games, his business will become very safe; the average gain of 20 cts per game is 460 $ 80 cts, or exactly 4 times as great as the mean error 2 $ 40 cts $\times$ 48 But who will enter upon such a game against the banker (a game, after all, which is not worse than so many others)? The very stake is already greater than the mathematical expectation, every prudent regard to part of the mean error will only augment the disproportion No prudent man will enter upon such a game, unless he can thereby avoid a greater risk in this way we insure our risks, because it is too dangerous to be "one's own insurer" If the game is arranged in such an entertaining way that we pay 40 cts for the excitement only of taking part in every game, then even rather a cautious person may also continue for 144 games, the mean error $(\pm 2{\cdot}4\sqrt{144}$ as above) being only 28 $ 80 cts or $144(0{\cdot}8 - (1{\cdot}0 - 0{\cdot}4))$ $ For a poor fellow, who has only one dollar in his pocket, but who must for some reason necessarily get 8 $, such a game may also be the best resource But if a man owns only 2304 $, and fails if he cannot get 8 times as much, then he would be exceedingly foolhardy if he played 2304 times or more in that bank If we must run the risk, we can do no better than venturing everything on one card, if we distribute our chances over n repetitions, then we must, beyond the mathematical expectation, hope for $\sqrt{n}$ times that part of the mean error which might help by the one attempt

Example 2. Two fire-insurance companies have each insured 10,000 farms for a total insurance of £ 10,000,000 The yearly probability of damage by fire is $\frac{1}{1000}$, and both must every year spend £ 5000 on management. Both have sufficient guaranty-fund to rest satisfied with one single mean error as security against a deficit in each fiscal year How high must either fix its annual premium, when there is the difference that the company A has 10,000 risks of £ 1,000, while B has insured

n		a	na	na^2
100	farms for	£ 10,000	£ 1,000,000	$10{,}000 \times (10)^6$
400	» »	» 5,000	» 2,000,000	10,000 »
1,500	» »	» 2,000	» 3,000,000	6,000 »
2,500	» »	» 1,000	» 2,500,000	2,500 »
2,000	» »	» 500	» 1,000,000	500 »
1,500	» »	» 200	300,000	60 »
2,000	» »	» 100	» 200,000	20 »
10,000	farms		£ 10,000,000	$29{,}080 \times (10)^6$

Since $p(1-p) = 0{\cdot}000999$, the mathematical expectation $\pm$ its mean error is in the case of A = £ 100,000 $\pm$ £ 3,161, in the case of B = £ 10,000 $\pm$ £ 5,390, the premiums are therefore £ 1 18s. 4d. and £ 2 7s. 10d. respectively for £ 1,000; i. e. B must reinsure part of its risks.

§ 76. The mean error and, in general, the law of error, of the total mathematical expectation for mutually bound events which may be considered co-ordinate events of the same trials, are computed in half-invariants by means of the sums of powers. If the trial can have n various events, of which the one whose probability is p_i entails a gain of the value a_i, and we imagine the same repeated a sufficiently large number of times (N times), the account will show

$$a_1 \text{ occurring } p_1 N \text{ times,}$$

$$a_n \text{ occurring } p_n N \text{ times.}$$

Hence

$$s_0 = (p_1 + \dots + p_n) N$$
$$s_1 = (p_1 a_1 + \dots + p_n a_n) N$$
$$s_2 = (p_1 a_1^2 + \dots + p_n a_n^2) N,$$

and the half-invariants for the single trial will be

$$\lambda_1 = p_1 a_1 + \dots + p_n a_n = \text{the total mathematical expectation} = H(1, \dots n);$$
$$\lambda_2 (H(1, \dots n)) = p_1 a_1^2 + \dots + p_n a_n^2 - (p_1 a_1 + \dots + p_n a_n)^2 \tag{144}$$

By this formula, therefore, we must in such cases compute the square of the mean error of the total mathematical expectation for the single trial. For the square of the mean error of the expectation from N trials we have consequently

$$\lambda_2 (N, H(1, \dots n)) = N(p_1 a_1^2 + \dots + p_n a_n^2 - H(1, \dots n)^2) \tag{145}$$

By even game we understand a game where the total mathematical expectation is 0, the last term of this formula will consequently disappear in such a game. As the mean error does not depend on the absolute values of the gains or losses, but only on

their differences, we may in the computation of the squares of the mean errors reduce to even game by subtracting the mathematical expectation from all the gains, and adding it to the losses. Thus we may write

$$\lambda_2(N, H(1, \ldots n)) = N(p_1(a_1 - H(1, \ldots n))^2 + \ldots + p_n(a_n - H(1, \ldots n))^2) \qquad (146)$$

This rule then differs from the rule of unbound games only in the absence of the factors $(1-p_1), \ldots (1-p_n)$.

We can now compute the mean errors in the examples 1 and 2, in § 74. In No. 1 we have

$$\lambda_2(H) = \tfrac{1}{2}(0)^2 + \tfrac{1}{4}(2)^2 + \tfrac{1}{4}(6)^2 - 2^2 =$$
$$= \tfrac{1}{2}(-2)^2 + \tfrac{1}{4}(0)^2 + \tfrac{1}{4}(4)^2 \quad = 6$$

In the life-annuity example we now see the advantage of using the longer formula (140) for the value of the annuity, rather than the formula (139) which gives the value as the sum of a number of endowments, for the partial expectations are here not unbound, and only the deaths in the several years of age exclude one another and can be considered co-ordinate events in the same game. For the square of the mean error of the life-annuity we have, from (144):

$$\frac{v^2}{\rho^2 l(a)} \sum_{x=1}^{x=\infty} (l(a+x) - l(a+x+1))(1-(1+\rho)^{-x})^2 -$$
$$- \frac{v^2}{\rho^2 l(a)^2} \left\{ \sum_{x=1}^{x=\infty} (l(a+x) - l(a+x+1))(1-(1+\rho)^{-x}) \right\}^2 \qquad (147)$$

§ 77. In the above studies on the mean errors of mathematical expectations we have supposed that the probabilities we use are free from error, being either determined a priori by good theory or found a posteriori from very large numbers of repetitions. This determination is not complete in the cases in which the probabilities determined a posteriori are found only by small numbers of trials, or if probabilities computed a priori presuppose values observed with sensibly large mean errors. The same warning must be taken with respect to other values which may enter into the computed mathematical expectations, the value of the gains, for instance, may depend on the future rate of interest. Whether some of the manifold sources of errors are to be omitted in a computation of the mean error, or not, must for each special case depend on the relative smallness of the parts of the total λ_2. As to the theory of probability it is characteristic only that the parts of the squares of the mean errors, considered in §§ 75 and 76, are, as a rule, very important, while the analogous parts in other problems are often insignificant. When the orbit of a planet is computed by the method of the least squares, then, in order to restrict the limits of research for its next discovery, we have to compute the mean errors of its co-ordinates

at the next opposition. Ordinarily these mean errors are so large that the λ_2 for its future observations may be wholly omitted, though this λ_2 is analogous to those from §§ 75 and 76. But when we have computed a table of mortality by the method of the least squares, we can certainly find by that method the mean error $\sqrt{\lambda_1(p)}$ of the probability of life computed from the table, but if we are to predict anything as to the uncertainty with regard to n lives, and with regard to the corresponding mathematical expectation npa, then we must not, unless n is very great, take the mean error as $na\sqrt{\lambda_1(p)}$, but we must, as a rule, first take $\lambda_1(H)$ in consideration, and consequently use the formula $a\sqrt{np(1-p)+n^2\lambda_1(p)}$. (Comp. example, § 72.)

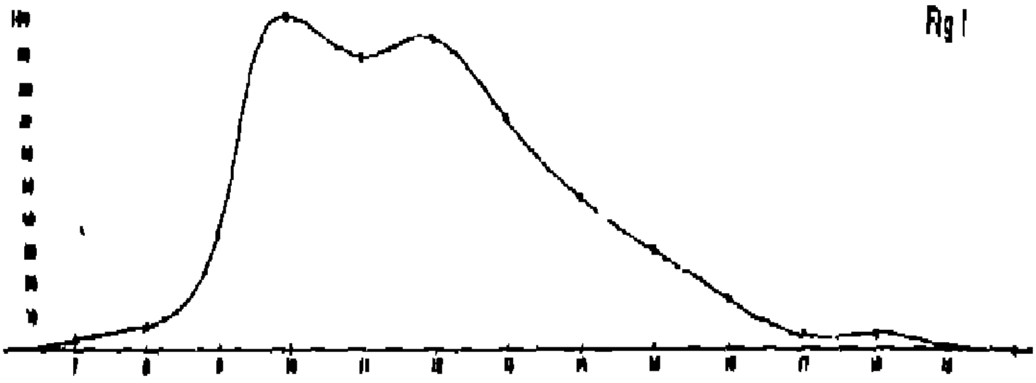


Fig 1

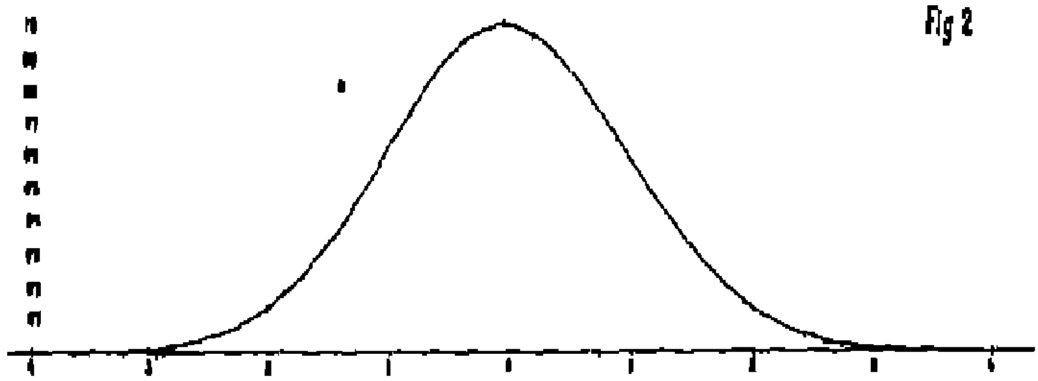


Fig 2

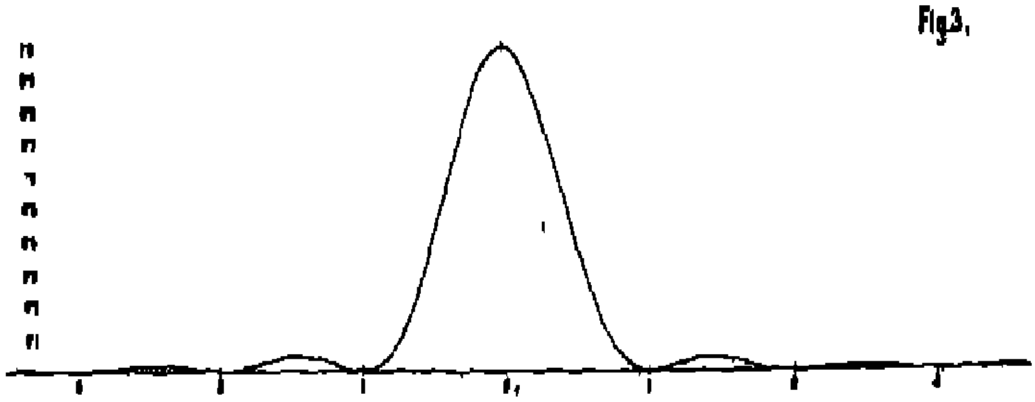


Fig 3.

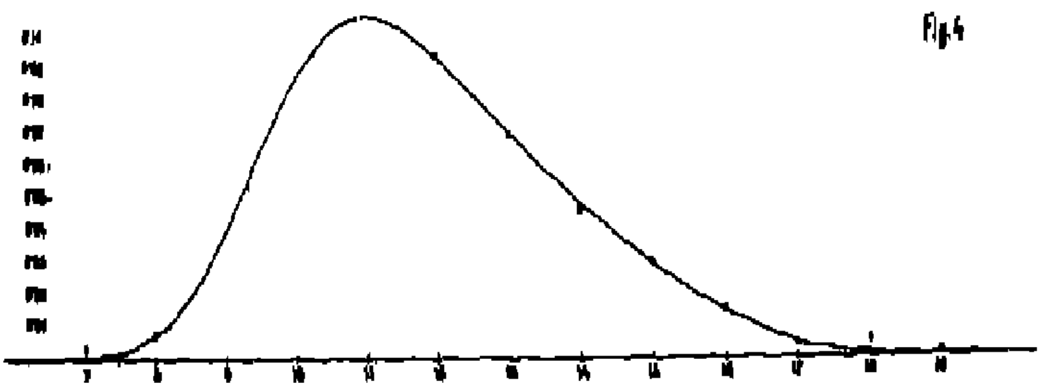


Fig 4

THE AMERICAN STATISTICAL ASSOCIATION

Vol. II No. 3

THE ANNALS

of

MATHEMATICAL STATISTICS

SIX DOLLARS PER ANNUM

AUGUST 1931

CORRECTION FOR THE MOMENTS OF A FREQUENCY DISTRIBUTION IN TWO VARIABLES[1]

By

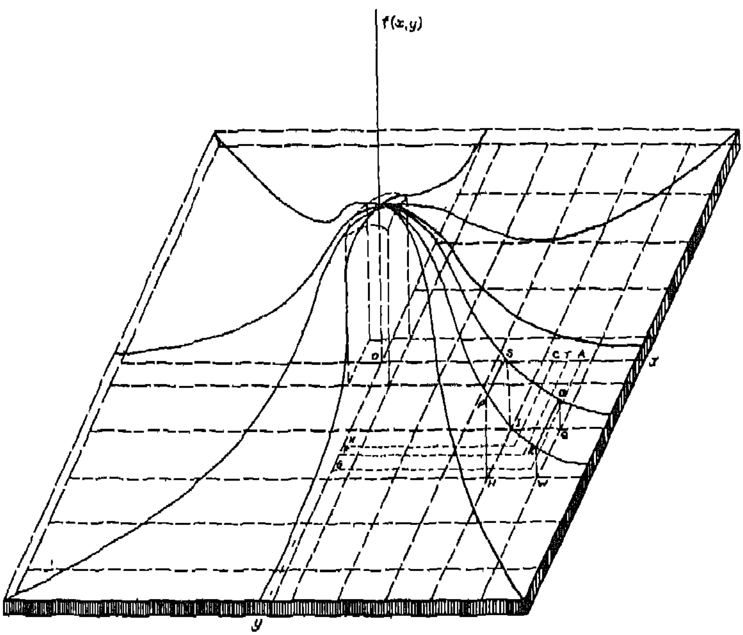
WILLIAM DOWELL BATEN
University of Michigan

In certain statistical problems it is beneficial to divide the given data into classes or groups and investigate the distribution in this form. The moments determined for the distribution divided into classes differ from the moments determined from the original data. It is the object of this article to show how to modify the former to secure the latter for a frequency distribution in two variables

After the data, given for a frequency distribution of one variable, have been divided into classes the class mark is then the representive of the items in a class. This is assuming that the mean of the items falling in a class is equal to the class mark. For a large number of items in a class, distributed throughout the entire class, the class mark differs very little from the average of the items in the class. But the average of the items raised to a power is not equal to the class mark raised to the same power. Hence corrections should be made to the moments determined from a distribution which is divided into classes.

For a distribution of two variables x and y the data are divided into xy-classes, where the class mark of an x-class

[1]Presented to the American Mathematical Society, Sept 12, 1930.



f(x,y)
O
S
C T A
x
y

is considered to be the representive of the items falling in this class, while the class mark of a y -class is the representive of all items in this particular class. The coordinates of the point in the xy -plane, whose abscissa is the class mark of the x -class and whose ordinate is the class mark of the y -class, may be considered to be the class mark of the double class or the xy class.

Let the frequencies of the distribution be represented by the volumes of the volume-compartments as shown in the figure. The sum of all such compartments is the total of the frequencies and should be equal to the number of items in the distribution. The little solid $HWQI\text{-}SPRD$ is the frequency of the items falling in the 5th x -class and in the 3rd y -class. OT and OF are the class marks of this x-class and this y -class. $(OT)^n(OF)^m$ multiplied by the frequency of the items falling in this double xy-class may differ considerably from the sum $(OC)^n(OK)^m + (OA)^n(OG)^m + \ldots$, hence corrections must be made to the moments obtained from the distribution divided into classes where the double class marks are the representives of the items in the class. If the class units are made smaller and are allowed to become very near to zero the errors are not so large, for it must be remembered that our results are only approximations.

By definition the n'th, m'th moment of the distribution which is divided into classes is

$$V'_{n;m} = \frac{\Sigma\Sigma x_i^n y_j^m}{N} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_i + h, y_j + k)\, dh\, dk,$$

where (x_i, y_j) is considered to be the class mark of the i, j -class, and the double summation extends over all the classes. It is further assumed that $f(x_i + h, y_j + k)$ is such a function which can be expanded into a Taylor series. The above becomes

$$\frac{\Sigma\Sigma\, x_i^n y_j^m}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_i+h, y_j+k)\,dh\,dk = \frac{\Sigma\Sigma\, x_i^n y_j^m}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}\Big\{f(x_i, y_j)$$

$$+h\frac{\partial f(x_i,y_j)}{\partial x_i}+k\frac{\partial f(x_i,y_j)}{\partial y_j}+\frac{1}{2!}\left[\frac{h^2\partial^2 f(x_i,y_j)}{\partial x_i^2}+\frac{2hk\,\partial^2 f(x_i,y_j)}{\partial x_i\,\partial y_j}\right.$$

$$\left.+\frac{k^2\partial^2 f(x_i,y_j)}{\partial y_j^2}\right]+\frac{1}{3!}\left[\frac{h^3\partial^3 f(x_i,y_j)}{\partial x_i^3}+\frac{3h^2k\,\partial^3 f(x_i,y_j)}{\partial x_i^2\partial y_j}+\frac{3hk^2\partial^3 f(x_i,y_j)}{\partial x_i\,\partial y_j^2}\right.$$

$$\left.+\frac{k^3\partial^3 f(x_i,y_j)}{\partial y_j^3}\right]+\cdots\Big\}dh\,dk = \frac{\Sigma\Sigma\, x_i^n y_j^m}{N}\Big\{f(x_i,y_j)$$

$$+\left[\frac{\partial^2 f(x_i,y_j)}{2^2 3!\,\partial x_i^2}+\frac{\partial^2 f(x_i,y_j)}{2^2 3!\,\partial y_j^2}\right]+\left[\frac{\partial^4 f(x_i,y_j)}{2^4 5!\,\partial x_i^4}+\frac{2\,\partial^4 f(x_i,y_j)}{2^4 3\, 5!\,\partial x_i^2\partial y_j^2}\right.$$

$$\left.+\frac{\partial^4 f(x_i,y_j)}{2^4 5!\,\partial y_j^4}\right]+\frac{1}{6!}\left[\frac{\partial^6 f(x_i,y_j)}{2^6\, 7\ \partial x_i^6}+\frac{\partial^6 f(x_i,y_j)}{2^6\,\partial x_i^4\,\partial y_j^2}\right.$$

$$\left.+\frac{\partial^6 f(x_i,y_j)}{2^6\,\partial x_i^2\,\partial y_j^4}+\frac{\partial^6 f(x_i,y_j)}{2^6\cdot 7\,\partial y_j^6}\right]+\cdots\Big\}.$$

Now use the Euler-Maclaurin Summation* formula for two variables for finding the value of this double summation. This formula is

*This formula is developed on pages 317-319.

$$\sum_c^d \sum_a^b U(x,y) = \int_c^{d+1}\int_a^{b+1} U(x,y)\,dx\,dy - \frac{1}{2}\int_c^{d+1} U(x,y)\,dy\Big]_a^{b+1} - \frac{1}{2}\int_a^{b+1} U(x,y)\,dx\Big]_a^{d+1}$$

$$+\frac{1}{12}\frac{\partial}{\partial y}\int_a^{b+1} U(x,y)\,dx\Big]_c^{d+1} + \frac{1}{12}\frac{\partial}{\partial x}\int_c^{d+1} U(x,y)\,dy\Big]_a^{b+1} + \frac{U(x,y)}{4}\Big]_c^{d+1}\Big]_a^{b+1} - \frac{\partial U(x,y)}{24\,\partial x}\Big]_c^{d+1}\Big]_a^{b+1}$$

$$-\frac{\partial U(x,y)}{24\,\partial y}\Big]_c^{d+1}\Big]_a^{b+1} - \frac{\partial^3}{720\,\partial x^3}\int_c^{d+1} U(x,y)\,dy\Big]_a^{b+1} - \frac{\partial^3}{720\,\partial y^3}\int_a^{b+1} U(x,y)\,dx\Big]_c^{d+1}$$

$$+\frac{\partial^2 U(x,y)}{144\,\partial x\,\partial y}\Big]_c^{d+1}\Big]_a^{b+1} + \frac{\partial^3 U(x,y)}{1440\,\partial x^3}\Big]_c^{d+1}\Big]_a^{b+1} + \frac{\partial^3 U(x,y)}{1440\,\partial y^3}\Big]_c^{d+1}\Big]_a^{b+1} + \cdots$$

which is the double summation of the function $U(x,y)$ from a to b on the x-axis and from c to d along the y-axis. Applying this formula to the double summation above

$$V'_{n,m} = \iint x^n y^m \Big\{ f(x,y) + \frac{1}{2!2^2}\left[\frac{\partial^2 f(x,y)}{3\,\partial x^2} + \frac{\partial^2 f(x,y)}{3\,\partial y^2}\right]$$

$$+\frac{1}{4!2^4}\left[\frac{\partial^4 f(x,y)}{5\,\partial x^4} + \frac{(\frac{4}{2})\,\partial^4 f(x,y)}{3\cdot 3\,\partial x^2\,\partial y^2} + \frac{\partial^4 f(x,y)}{5\,\partial y^4}\right]$$

$$+\frac{1}{6!2^6}\left[\frac{\partial^6 f(x,y)}{7\,\partial x^6}+\frac{\binom{6}{2}\partial^6 f(x,y)}{5\cdot 3\,\partial x^4\partial y^2}+\frac{\binom{6}{4}\partial^6 f(x,y)}{3\;5\,\partial x^2\partial y^4}+\frac{\partial^6 f(x,y)}{7\,\partial y^6}\right]$$

$$+\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+\frac{1}{s!2^s}\left[\frac{\partial^s f(x,y)}{(s+1)\partial x^s}+\frac{\binom{s}{2}\partial^s f(x,y)}{(s-2+1)(2+1)\partial x^{s-2}\partial y^2}+\frac{\binom{s}{4}\partial^s f(x,y)}{(s-4+1)(4+1)\,\partial x^{s-4}\partial y^4}\right.$$

$$+\frac{\binom{s}{6}\partial^s f(x,y)}{(s-6+1)(6+1)\,\partial x^{s-6}\partial y^6}+\cdots+\frac{\binom{s}{t}\partial^s f(x,y)}{(s-t+1)(t+1)\,\partial x^{s-t}\,\partial y^t}$$

$$\left.\left.+\quad\cdots\cdots\cdots+\frac{\partial^s f(x,y)}{(s+1)\,\partial y^s}\right]\right\}dx\,dy+0+0+\cdots;$$

t is an even number. In obtaining this result it was assumed that $f(x,y)$, $f'(x,y)$, $x^k y^w f(x,y)$, $x^k y^w f'(x,y)$ vanish or become negligible at the limits on the x and y axes, k and w are positive integers.

Therefore

$$V_{n\,m}=\mu'_{n\,m}+\frac{2!}{2^2 3!}\binom{n}{2}\mu'_{n-2\cdot m}+\frac{2!}{2^2 3!}\binom{m}{2}\mu'_{n\cdot m-2}$$

$$+\frac{1}{4!2^4}\left[\frac{4!}{5}\binom{n}{4}\mu'_{n-4:m}+\frac{(2!)^2}{3\cdot 3}\binom{4}{4}\binom{n}{2}\binom{m}{2}\mu'_{n-2:m-2}+\frac{4!}{5}\binom{m}{4}\mu'_{n:m-4}\right.$$

$$+\frac{1}{6!2^6}\left[\frac{6!}{7}\binom{n}{6}\mu_{n-4:m}+\frac{4!2!}{5\;3}\binom{6}{2}\binom{n}{4}\binom{m}{2}\mu'_{n-4\cdot m-2}+\frac{2!4!}{3\;5}\binom{6}{4}\binom{n}{2}\binom{m}{4}\mu_{n-2.m-4}\right.$$

$$\left.+\frac{6!}{7}\binom{m}{6}\mu'_{n.m-6}\right]+\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+\frac{1}{s!2^s}\left[\frac{s!}{(s+1)}\binom{n}{s}\mu'_{n-s\cdot m}+\frac{(s-2)!\,2!}{(s-2+1)(2+1)}\binom{s}{2}\binom{n}{s-2}\binom{m}{2}\mu'_{n-\overline{s-2}:m-2}\right.$$

$$+\frac{(s-4)!4!}{(s-4+1)(5)}\binom{s}{4}\binom{n}{s-4}\binom{m}{4}\mu'_{n-\overline{s-4}:m-4}$$

$$+\frac{(s-6)!6!}{(s-6+1)(7)}\binom{s}{6}\binom{n}{s-6}\binom{m}{6}\mu'_{n-\overline{s-6}\cdot m-6}+\cdots$$

$$\cdots+\frac{(s-t)!t!}{(s-t+1)(t+1)}\binom{s}{t}\binom{n}{s-t}\binom{m}{t}\mu'_{n-\overline{s-t}:m-t}+\cdots$$

$$\left.+\frac{s!}{(s-1)}\binom{m}{s}\mu'_{n:m-s}\right]+\cdots\cdots$$

If $m=0$ the formula becomes the formula for obtaining the moments about a fixed origin for one variable This has been done by Sheppard and Carver.

If n and m take on integral values

$$V'_{1\cdot 1}=\mu'_{1\cdot 1},$$

$$V'_{2\,1}=\mu'_{2.1}+\frac{1}{12}\mu'_{0\cdot 1},\; V'_{1\cdot 2}=\mu'_{1\cdot 2}+\frac{1}{12}\mu'_{1:0},$$

$$V'_{2\cdot 2}=\mu'_{2.2}+\frac{1}{12}\mu'_{0:2}+\frac{1}{12}\mu'_{2\cdot 0}+\frac{1}{144},$$

$$V'_{3\cdot 1}=\mu'_{3:1}+\frac{1}{4}\mu'_{1:1},\; V'_{1:3}=\mu'_{1\,3}+\frac{1}{4}\mu'_{1\,1},$$

$$V'_{3:2}=\mu'_{3\cdot 2}+\frac{1}{4}\mu'_{1\cdot 2}+\frac{1}{12}\mu'_{3:0}+\frac{1}{48}\mu'_{1:0},$$

$$V'_{2\,3}=\mu'_{2\,3}+\frac{1}{12}\mu'_{0\,3}+\frac{1}{4}\mu'_{2:1}+\frac{1}{48}\mu'_{0\cdot 1},$$

$$V_{3:3}=\mu'_{3\cdot 3}+\frac{1}{4}\mu'_{1\,3}+\frac{1}{4}\mu'_{3\cdot 1}+\frac{1}{16}\mu'_{1:1},$$

$$V'_{4\,1}=\mu'_{4\cdot 1}+\frac{1}{2}\mu'_{2:1}+\frac{1}{80}\mu'_{0:1},\; V'_{1:4}=\mu'_{1:4}+\frac{1}{2}\mu'_{1\cdot 2}+\frac{1}{80}\mu'_{1:0},$$

$$V'_{4:2}=\mu'_{4:2}+\frac{1}{2}\mu'_{2:2}+\frac{1}{12}\mu'_{4:0}+\frac{1}{80}\mu'_{0:2}+\frac{1}{24}\mu'_{2\cdot 0}+\frac{1}{960}.$$

$$\nu'_{2:4} = \mu'_{2:4} + \frac{1}{12}\mu'_{0:4} + \frac{1}{2}\mu'_{2:2} + \frac{1}{80}\mu'_{2:0} + \frac{1}{24}\mu'_{0:2} + \frac{1}{960},$$

$$\nu'_{4:3} = \mu'_{4:3} + \frac{1}{2}\mu'_{2:3} + \frac{1}{4}\mu'_{4:1} + \frac{1}{80}\mu'_{0:3} + \frac{1}{8}\mu'_{2:1} + \frac{1}{260}\mu'_{0:1},$$

$$\nu'_{3:4} = \mu'_{3:4} + \frac{1}{2}\mu'_{3:2} + \frac{1}{4}\mu'_{1:4} + \frac{1}{80}\mu'_{3:0} + \frac{1}{8}\mu'_{1:2} + \frac{1}{260}\mu_{1:0},$$

$$\nu'_{4:4} = \mu'_{4:4} + \frac{1}{2}\mu'_{2:4} + \frac{1}{2}\mu'_{4:2} + \frac{1}{80}\mu'_{0:4} + \frac{1}{4}\mu'_{2:2} + \frac{1}{80}\mu'_{4:0}$$

$$+ \frac{1}{160}\mu'_{2:0} + \frac{1}{160}\mu'_{0:2} + \frac{1}{6400},$$

. . . .

From the above the μ''s can be obtained.

$$\mu'_{1:1} = \nu'_{1:1},$$

$$\mu'_{2:1} = \nu'_{2:1} - \frac{1}{12}M_y, \quad \mu'_{1:2} = \nu'_{1:2} - \frac{1}{12}M_x,$$

$$\mu'_{3:1} = \nu'_{3:1} - \frac{1}{4}\nu'_{1:1}, \quad \mu'_{1:3} = \nu'_{1:3} - \frac{1}{4}\nu'_{1:1},$$

$$\mu'_{3:2} = \nu'_{3:2} - \frac{1}{4}\nu'_{2:1} - \frac{1}{12}\nu'_{3:0}, \quad \mu'_{2:3} = \nu'_{2:3} - \frac{1}{4}\nu'_{1:2} - \frac{1}{12}\nu'_{0:3},$$

$$\mu'_{3:3} = \nu'_{3:3} - \frac{1}{4}\nu'_{3:1} - \frac{1}{4}\nu'_{1:3} + \frac{1}{16}\nu'_{1:1}, \quad \text{etc.}$$

By translating the origin to (M_x, M_y)

$$\mu_{1:1}=\nu_{1:1},$$

$$\mu_{2\cdot 1}=\nu'_{2:1}, \quad \mu_{1.2}=\nu_{1:2},$$

$$\mu_{22}=\nu_{2\cdot 2}-\frac{1}{12}(\nu_{0\cdot 2}+\nu'_{2.0})+\frac{1}{144},$$

$$\mu_{3\cdot 1}=\nu_{3:1}-\nu_{1:1}, \quad \mu_{1.3}=\nu_{1:3}-\nu_{1:1},$$

$$\mu_{3.2}=\nu_{3:2}-\frac{1}{2}\nu_{2\,1}-\frac{1}{12}\nu_{3\cdot 0},$$

$$\mu_{2.3}=\nu_{2:3}-\frac{1}{2}\nu_{1:2}-\frac{1}{12}\nu_{0.3},$$

$$\mu_{3\cdot 3}=\nu'_{3.3}-\frac{1}{4}\nu_{1:3}-\frac{1}{4}\nu'_{3\cdot 1}+\frac{1}{16}\nu_{1:1},$$

etc.

In making corrections for the double moments it must be remembered to correct the single moments of the x's and the y's.*

Euler-Maclaurin Summation for Two Variables

Suppose it is possible to find a function $g(x,y)$ such that $g(x+1, y+1)-g(x+1,y)-g(x,y+1)+g(x,y)= f(x,y)$, or $\underset{x}{\Delta}\underset{y}{\Delta}\, g(x,y)=f(x,y)$ or $\underset{x}{\Delta}^{-1}\underset{y}{\Delta}^{-1} f(x,y) = g(x,y)$, where Δ represents finite difference and Δ^{-1} represents finite integration. If $g(x,y)$ is such a function, then

$$g(a+1,c+1)-g(a+1,\;c)-g(a\;,c+1)+g(a\;,\;c)=f(a\;,\;c),$$

$$g(a+2,c+1)-g(a+2,\;c)-g(a+1,c+1)+g(a+1,\;c)=f(a+1,\;c),$$

$$g(a+1,c+2)-g(a+1,c+1)-g(a\;,c+2)+g(a\;,c+1)=f(a\;,c+1),$$

$$g(a+2,c+2)-g(a+2,c+1)-g(a+1,c+2)+g(a+1,c+1)=f(a+1,c+1),$$

---

*See Frequency Curves by H. C Carver in Handbook of Math. Statistics.

. . . .

. . . .

$$g(b,d) \quad -g(b\ ,d-1)-g(b-1,\ d)+g(b-1,d-1)=f(b-1,c-1),$$

$$\underline{g(b+1,d+1)-g(b+1,\ d)-g(b\ ,d+1)+g(b,d)=f(b,d).}$$

Add: $g(b+1,d+1)-g(b+1,\ c)-g(a\ ,d+1)+g(a,c)=\sum_{c}^{d}\sum_{a}^{b} f(x,y).$

Or
$$\sum_{c}^{d}\sum_{a}^{b} f(x,y)=g(x,y)\Big]_{c}^{d+1}\Big]_{a}^{b+1}$$

If it is possible to find the function $g(x,y)$ then the double sum $\sum_{c}^{d}\sum_{a}^{b} f(x,y)$ can be found. Expand $g(x+1,y+1)$ in a Taylor series.

$$g(x+1,y+1)=g(x,y)+\frac{\partial g}{\partial x}+\frac{\partial g}{\partial y}+\frac{1}{2!}\left[\frac{\partial^2 g}{\partial x^2}+\frac{2\partial^2 g}{\partial x\,\partial y}+\frac{\partial^2 g}{\partial y^2}\right]$$

$$+\frac{1}{3!}\left[\frac{\partial^3 g}{\partial x^3}+\frac{3\partial^3 g}{\partial x^2\partial y}+\frac{3\partial^3 g}{\partial x\partial y^2}+\frac{\partial^3 g}{\partial y^3}\right]+\cdot$$

$$=(e^{\frac{\partial}{\partial x}+\frac{\partial}{\partial y}}\,g(x,y)=(e^{\mathcal{D}+\mathcal{D}})\,g(x,y),$$

where $\mathcal{D}, \mathcal{D}^h, \mathcal{D}^r\mathcal{D}^s$ represent respectively

$$\frac{\partial}{\partial x}g(x,y),\ \frac{\partial^h}{\partial x^h}g(x,y),\ \frac{\partial^{r+s}}{\partial x^r\,\partial y^s}g(x,y).$$

Hence

$$g(x+1,y+1)-g(x+1,y)-g(x,y+1)+g(x,y)$$
$$=(e^{\mathcal{D}+\mathcal{D}}-e^{\mathcal{D}}-e^{\mathcal{D}}+1)\,g(x,y)$$
$$=\{(e^{\mathcal{D}}-1)(e^{\mathcal{D}}-1)\}\,g(x,y).$$

where the D's are operators operating on the function $g(x,y)$. Therefore

$$g(x,y)=\frac{1}{[(e^{\mathcal{D}}-1)(e^{\mathcal{D}}-1)]}\,f(x,y),$$

where the operators are now operating upon the function $f(x,y)$.

To develop $\dfrac{1}{(e^{u}-1)(e^{v}-1)}$ into a Taylor series it is necessary to develop $\dfrac{uv}{(e^{u}-1)(e^{v}-1)}$ into a Taylor series and then divide by uv. This becomes after D_x, D_y replace u and v respectively,

$$\left\{\frac{1}{[(e^{D_x}-1)(e^{D_y}-1)]}\right\} f(x,y)=\left\{\frac{1}{D_xD_y}-\frac{1}{2D_y}-\frac{1}{2D_x}+\frac{1}{2!}\frac{D_x}{6D_y}+\frac{1}{2}+\frac{D_y}{6D_x}\right.$$

$$-\frac{D_x}{24}-\frac{D_y}{24}-\frac{D_x^3}{720D_y}+\frac{D_xD_y}{144}-\frac{D_y^3}{720D_x}+\frac{D_x^3}{1440}+\frac{D_y^3}{1440}$$

$$\left.+\frac{D_x^5}{6!42D_y}-\frac{D_xD_y^3}{6!12}-\frac{D_x^3D_y}{6!12}+\frac{D_y^5}{6!42D_x}\cdots\right\} f(x,y),$$

where $\dfrac{1}{D_x}$, $\dfrac{1}{D_y}$ represent integration.

Using these results $\displaystyle\sum_c^d\sum_a^b f(x,y)=g(x,y)\Big]_c^{d+1}\Big]_a^{b+1}$ or

$$\sum_c^d\sum_a^b f(x,y)=\int_c^{d+1}\int_a^{b+1} f(x,y)\,dx\,dy-\frac{1}{2}\int_c^{d+1} f(x,y)\,dy\Big]_a^{b+1}-\frac{1}{2}\int_a^{b+1} f(x,y)\,dx\Big]_c^{d+1}$$

$$+\frac{\partial}{12\,\partial y}\int_a^{b+1} f(x,y)\,dx\Big]_c^{d+1}+\frac{1}{12}\cdot\frac{\partial}{\partial x}\int_c^{d+1} f(x,y)\,dy\Big]_a^{b+1}+\frac{f(x,y)}{4}\Big]_c^{d+1}\Big]_a^{b+1}$$

$$-\frac{\partial f(x,y)}{24\,\partial x}\Big]_c^{d+1}\Big]_a^{b+1}-\frac{\partial f(x,y)}{24\,\partial y}\Big]_c^{d+1}\Big]_a^{b+1}-\frac{\partial^3}{720\,\partial x^3}\int_c^{d+1} f(x,y)\,dy\Big]_a^{b+1}$$

$$+\frac{\partial^2 f(x,y)}{144\,\partial x\,\partial y}\Big]_c^{d+1}\Big]_a^{b+1}-\frac{\partial^3}{720\,\partial y^3}\int_a^{b+1} f(x,y)\,dx\Big]_c^{d+1}$$

$$+\frac{\partial^3 f(x,y)}{1440\,\partial x^3}\Big]_c^{d+1}\Big]_a^{b+1}+\frac{\partial^3 f(x,y)}{1440\,\partial y^3}\Big]_c^{d+1}\Big]_a^{b+1}\cdots\cdots$$

W. D. Baten

THE STANDARD ERROR OF A MULTIPLE REGRESSION EQUATION[1]

By

JOHN RICE MINER

Since a multiple regression equation is essentially a hyperplane, fitted by the method of least squares, its standard error may be obtained from Gauss' *standard error of a function* recently discussed by Schultz (1930). Let the equation be

$$x_1 = b_{12.34\cdots m}x_2 + b_{13.24\cdots m}x_3 + \cdots + b_{1m.23\cdots(m-1)}x_m$$

where x_1 is the dependent variable, $x_2, x_3 \cdots x_m$ the independent variables, each measured from its respective mean, and $b_{12.34\cdots m}, \cdots, b_{1m.23\cdots(m-1)}$ the partial regression coefficients. Then the determinant of Schultz's equation (10) becomes

$$D = \begin{vmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & \Sigma x_2^2 & \Sigma x_2 x_3 & \cdot & \Sigma x_2 x_m \\ 0 & \Sigma x_2 x_3 & \Sigma x_3^2 & \cdot & \Sigma x_3 x_m \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \Sigma x_2 x_m & \Sigma x_3 x_m & \cdot & \Sigma x_m^2 \end{vmatrix} = n^m \sigma_2^2 \sigma_3^2 \cdot \ \ \sigma_m^2$$

(1)

$$\begin{vmatrix} 1 & 0 & 0 & & 0 \\ 0 & 1 & r_{23} & \cdot & r_{2m} \\ 0 & r_{23} & 1 & \cdot & r_{3m} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & r_{2m} & r_{3m} & \cdots & 1 \end{vmatrix} = n^m \sigma_2^2 \sigma_3^2 \cdots \sigma_m^2 \Delta, \text{ say}$$

[1]From the Department of Biology of the School of Hygiene and Public Health of the Johns Hopkins University.

Let Δ_{ij} be the cofactor of the element in the i 'th row and the j 'th column. Then

$$[\alpha\alpha] = D_{11}/D = \frac{1}{n},$$

$$[\beta\beta] = D_{22}/D = \Delta_{22}/n\sigma_2^2\Delta,$$

. . . .

$$[\mu\mu] = D_{mm}/D = \Delta_{mm}/n\sigma_m^2\Delta,$$

$$[\alpha\beta] = D_{12}/D = 0,$$

. . . .

$$[\alpha\mu] = D_{1m}/D = 0,$$

$$[\beta\gamma] = D_{23}/D = \Delta_{23}/n\sigma_2\sigma_3\Delta,$$

. . . .

$$[\beta\mu] = D_{2m}/D = \Delta_{2m}/n\sigma_2\sigma_m\Delta,$$

. . . .

$$\frac{\partial f}{\partial A} = 1, \; \frac{\partial f}{\partial B} = x_2, \; \frac{\partial f}{\partial C} = x_3, \cdots, \frac{\partial f}{\partial M} = x_m, \qquad \text{and}$$

$$\varepsilon^2 = \frac{n}{n-m}\sigma_1^2\left(1 - R^2_{1(23\cdots m)}\right).$$

Therefore, substituting these values in Schultz's equation (27.1), we have

$$\sigma_f = \frac{\sigma_1}{(n-m)^{\frac{1}{2}}}\left(1-R^2_{1(23\ldots m)}\right)^{\frac{1}{2}}\left\{1+\frac{\Delta_{22}}{\sigma_2^2\Delta}x_2^2+\frac{\Delta_{33}}{\sigma_3^2\Delta}x_3^2+\ldots+\frac{\Delta_{mm}}{\sigma_m^2\Delta}x_m^2\right.$$

(2)

$$\left.+2\frac{\Delta_{23}}{\sigma_2\sigma_3\Delta}x_2x_3+\ldots+2\frac{\Delta_{2m}}{\sigma_2\sigma_m\Delta}x_2x_m+\ldots\right\}^{\frac{1}{2}}.$$

For a simple regression equation this reduces to

$$(3) \quad \sigma_f = \frac{\sigma_1}{(n-2)^{\frac{1}{2}}}\left(1-r_{12}^2\right)^{\frac{1}{2}}\left\{1+\frac{x_2^2}{\sigma_2^2}\right\}^{\frac{1}{2}}.$$

This agrees with the expression given by Pearson (1913), if we remember that x_2 is measured from its mean and that Pearson does not correct for the number of parameters.

For a regression equation with two independent variables

$$\sigma_f = \frac{\sigma_1}{(n-3)^{\frac{1}{2}}}\left(1-R^2_{1(23)}\right)^{\frac{1}{2}}$$

$$(4) \quad \left\{1+\frac{x_2^2}{\sigma_2^2(1-r_{23}^2)}+\frac{x_3^2}{\sigma_3^2(1-r_{23}^2)}-\frac{2r_{23}x_2x_3}{\sigma_2\sigma_3(1-r_{23}^2)}\right\}^{\frac{1}{2}}$$

$$= \frac{\sigma_{1.23}}{(n-3)^{\frac{1}{2}}}\left\{1+\frac{x_2^2}{\sigma_{2.3}^2}+\frac{x_3^2}{\sigma_{3.2}^2}-\frac{2r_{23}x_2x_3}{\sigma_{2.3}\sigma_{3.2}}\right\}^{\frac{1}{2}}$$

As an example of the application of this formula we may calculate the standard error of the mean heart-weight (X_1) of the array of persons with an age (X_2) of 52 92 years and a

body-weight (X_3) of 49.93 kilograms in a population of 213 persons characterized by the following biometric constants:

$$M_1 = 348.9 \text{ g}; \qquad \sigma_1 = 79.4 \text{ g}; \qquad r_{12} = +0.114$$
$$M_2 = 59.65 \text{ yrs}; \qquad \sigma_2 = 17.54 \text{ yrs.}; \qquad r_{13} = +0.652$$
$$M_3 = 56.45 \text{ kg}; \qquad \sigma_3 = 14.38 \text{ kg}; \qquad r_{23} = -0.185.$$

From these data $r_{12.3} = +0.315$ and $r_{13.2} = +0.689$ and the regression equation of heart-weight on age and body-weight is

$$X_1 = 66.09 + 1.100 X_2 + 3.848 X_3$$

from which the mean heart-weight of persons aged 52.92 years and weighing 49.93 kg. is 316.4 g.

Substituting the appropriate values of the constants in (4) and remembering that $x_2 = X_2 - M_2 = -6.72$, $x_3 = X_3 - M_3 = -6.52$, and

$$(1 - R^2_{1(23)})^{\frac{1}{2}} = (1 - r^2_{12})^{\frac{1}{2}} (1 - r^2_{13.2})^{\frac{1}{2}},$$

$$\sigma_f = \frac{79.4}{210^{\frac{1}{2}}} (0.993)(0.725) \left\{ 1 + \frac{(-6.72)^2}{(17.54)^2 (0.966)} + \frac{(-6.52)^2}{(14.38)^2 (0.966)} \right.$$

$$\left. - \frac{2(-0.185)(-6.72)(-6.52)}{(17.54)(14.38)(0.966)} \right\}^{\frac{1}{2}} = 4.7 \text{ g}.$$

John Rice Miner

REFERENCES

Pearson, Karl. 1913. On the probable errors of frequency constants. Biometrika, 9:1-10

Schultz, Henry. 1930. The standard error of a forecast from a curve, J. Amer Stat Assoc., 25:139-185.

SAMPLING IN THE CASE OF CORRELATED OBSERVATIONS

By

CECIL C. CRAIG

National Research Fellow

Dr E C. Rhodes, in a paper in the *Journal of the Royal Statistical Society*,[1] has considered the distribution of characteristics of samples of N when the individual observations are not assumed to be independent. As he points out, there are many important cases in which the usual assumption of independence or randomness in the observations is not justifiable. In the present paper will be explained a method based on the semi-invariants of Thiele for the calculation of the characteristics of the sought distributions in this case which is especially to be preferred to the method based on moments when it is supposed that the observations are normally correlated. In the case it is further assumed that only consecutive observations are correlated, in addition to Dr. Rhodes' results, the third semi-invariant (which is the same as the third moment about the mean) of the variance and the mean and the variance of the third and fourth moments about the mean are given.

Let the N observations composing a sample be given by values of x_1, x_2, $\cdots$, x_N respectively and let $F_N(x_1, x_2, \cdots, x_N)$ be the n-way probability function of $x_1, x_2, \cdots$, and x_N.

[1]The Precision of Means and Standard Deviations When the Individual Errors Are Correlated, Vol. 90 (1927), pp. 135-143.

Then the semi-invariants, $\lambda_{rst}\cdots$ of $x_1, x_2, \cdots, x_N$ are defined by

$$e^{\left(\sum_1^N \lambda_i t_i\right)+\frac{1}{2}\left(\sum_1^N \lambda_i t_i\right)^{(2)}+\frac{1}{3!}\left(\sum_1^N \lambda_i t_i\right)^{(3)}+\cdots} \tag{1}$$
$$=\int_{-\infty,\cdots,-\infty}^{\infty,\cdots,\infty} dF_N(x_1, x_2, \cdots, x_N)\, e^{\left(\sum_1^N x_i t_i\right)}$$ [1]

which is to be regarded as a formal identity in $t_1, t_2, \cdots, t_N$. $\left(\sum_1^N \lambda_i t_i\right)^{(K)}$ is first expanded by the multinomial law and then each term $\lambda_1^r, \lambda_2^s, \lambda_3^t \cdots$ in the result is replaced by $\lambda_{rst}\cdots$.

We shall pass over the characteristics of distributions of means, since the method of semi-invariants is equivalent to that of moments in this case, and take up the distribution of moments about the mean in samples of N. Following the method previously used by the author in the case of independent observations,[2] let

$$\delta_i = x_i - \sum_1^N \frac{x_i}{N} \tag{2}$$
$$= \sum_{j=1}^N a_{ij} x_j \quad \text{with} \quad \begin{cases} a_{ij} = -\frac{1}{N} \\ a_{ii} = \frac{N-1}{N} \end{cases}$$

Then let $V(\delta_1, \delta_2, \cdots, \delta_{N-1})$ be the probability function of $\delta_1, \delta_2, \cdots, \delta_{N-1}$, $(\sum_i \delta_i = 0$. The semi-invariants $\lambda'_{rst}\cdots$ of $\delta_1, \delta_2, \cdots, \delta_{N-1}$ are defined by

[1]Following Cramér, I distinguish between probability and frequency functions. $F_N(x_1, x_2, \cdots, x_N)$ is the "cumulative" frequency function and thus the integral is an n-way Stieltjes integral.

[2]An Application of Thiele's Semi-invariants to the Sampling Problem; Metron, Vol. 7, No 4 (1928), pp. 3-75.

$$e^{\left(\sum_1^{N-1}\lambda_i' t_i\right)+\frac{1}{2}\left(\sum_1^{N-1}\lambda_i' t_i\right)^{(2)}+\frac{1}{3!}\left(\sum_1^{N-1}\lambda_i' t_i\right)^{(3)}+\cdots}$$

$$(3)\qquad =\int_{-\infty,\ \cdots,\ -\infty}^{\infty,\ \cdots,\ \infty} dV(\delta_1,\ \cdots,\ \delta_{N-1})\, e^{\left(\sum_1^{N-1}\delta_i t_i\right)}$$

$$=\int_{-\infty,\ \cdots,\ -\infty}^{\infty,\ \cdots,\ \infty} dF_N(x_1,\, x_2,\, \cdots,\, x_N)\, e^{\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N} a_{ij}\, x_j\, t_i\right)}$$

We have at once,

$$\left(\sum_{i=1}^{N-1} t_i \sum_{j=1}^{N} \lambda_j a_{ij}\right)^{(K)} = \left(\sum_{i=1}^{N-1}\lambda_i' t_i\right)^{(K)}$$

and as the author has previously remarked,[1] we can also write

$$(4)\qquad \left(\sum_1^N t_i \sum_{j=1}^{N} \lambda_j a_{ij}\right)^{(K)} = \left(\sum_1^N \lambda_i' t_i\right)^{(K)}$$

so long as the relation is only used to find the values of $\lambda'_{rst\cdots}$'s in which at least one of the subscripts is zero.

Then $S_K(\nu_n)$, the k 'th semi-invariant of the n 'th moment about the mean in samples of N, is given by the formula

$$S_K(\nu_n)=$$

$$\frac{1}{N^K}\Sigma\Sigma\cdots\frac{(-1)^{(r+s+t+\cdots)-1}\left[(r+s+t+\cdots)-1\right]!\,K!\,\nu'^{\,r}_{a_1 n,\, a_2 n,\cdots}\,\nu'^{\,s}_{b_1 n,\, b_2 n,\cdots}\,\nu'^{\,t}_{c_1 n,\, c_2 n,\cdots}\cdots}{\left[a_1!\, a_2!\cdots\right]^r\left[b_1!\, b_2!\cdots\right]^s\left[c_1!\, c_2!\cdots\right]^t\, r!\, s!\, t!\cdots}$$

---

[1]loc. cit., pp 18, 19.

the notation $\nu'_{u\,v\,w}$ referring to moments of $\delta_1, \cdots \delta_{N-1}, \delta_N$, the summation including all terms for which

$$r(a_1 + a_2 + \cdots) + s(b_1 + b_2 + \cdots) + t\,(c_1 + c_2 + \cdots) + \cdots = k,$$

$$a_1 \geq a_2 \geq \cdots\cdots$$

$$b_1 \geq b_2 \geq \cdots$$

$$c_1 \geq c_2 \geq \cdots$$

$$\cdots\quad\cdots$$

$$(a_1 + a_2 + \cdots) > (b_1 + b_2 + \cdots) > (c_1 + c_2 + \cdots) > \cdots\cdots .$$

In particular:

$$S_1(\nu_n) = \frac{1}{N}\Sigma\,\nu'_{n,0}\,, \quad \left(\Sigma\nu'_{n,0} = \nu'_{n,0\cdots0} + \nu'_{0,n,0\cdots0} + \nu'_{0,0,n,0\cdots0} + \cdots\right),$$

$$S_2(\nu_n) = \frac{1}{N^2}\left[\Sigma\,\nu'_{2n,0} + 2\Sigma\,\nu'_{n,n,0} - (\Sigma\,\nu'_{n,0})^2\right],$$

(5)

$$S_3(\nu_n) = \frac{1}{N^3}\Big[\Sigma\,\nu'_{3n,0} + 3\Sigma\,\nu'_{2n,n,0} + 6\Sigma\,\nu'_{n,n,n,0} - 3(\Sigma\,\nu'_{2n,0})(\Sigma\,\nu'_{n,0})$$

$$- 6(\Sigma\,\nu'_{n,n,0})(\Sigma\,\nu'_{n,0}) + 2(\Sigma\,\nu'_{n,0})^3\Big] \quad ,$$

$$\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots$$

On writing out the moments $\nu'_{u\,v\,w}$ in terms of the semi-invariants λ'_{rst} ·[2] and then using (4) the sought semi-invariants are obtained

In the case that the N observations are normally correlated and $F_N(x_1, x_2, \cdots, x_N)$ is the N -dimensional normal probability function, the left-hand member of (4) vanishes for $k \geq 3$.

If we suppose that the standard deviations of $x_1, x_2, \cdots, x_N$ are all equal (which we shall always do) and take as the simplest case that $x_1, x_2, \cdots, x_N$ are normally correlated and that

---

[1]See the author's paper cited, p. 21, formula (25).

[2]For a detailed explanation of this kind of calculation see the author's paper cited, pp. 23-27.

the correlation as measured by the Pearsonian coefficient, $r_{x_i x_j}$, is the same for each pair, x_i, x_j, of the set of N observations, we get

$$\lambda'_{20} = \lambda'_{020} = \lambda'_{0020} = \cdots = \frac{N-1}{N}(\lambda_{20} - \lambda_{11}) = \frac{N-1}{N}(1-r)\lambda_{20},$$

$$\lambda'_{110} = \lambda'_{1010} = \lambda'_{0110} = \cdots\cdots = -\frac{1}{N}(\lambda_{20} - \lambda_{11}) = -\frac{1}{N}(1-r)\lambda_{20},$$

if the common value of $r_{x_i x_j}$ be denoted simply by r. But if the observations are independent and the parent population is normal we have

$$\lambda'_{20} = \lambda'_{020} = \lambda'_{0020} = \cdots = \frac{N-1}{N}\lambda_{20},$$

$$\lambda'_{110} = \lambda'_{1010} = \lambda'_{0110} = \cdots = -\frac{1}{N}\lambda_{20}.$$

Thus it follows that the distributions of the characteristics of samples of N in this particular case of dependent observations are the same as if the observations were independent and taken from a normal population of variance $(1-r)\lambda_{20}$.

In case $F_N(x_1, x_2, \cdots, x_N)$ is normal it is convenient to express the right hand members of (5) directly in terms of the semi-invariants $\lambda'_{rst\ldots}$ for $n = 2, 3, 4$. For that purpose we shall adopt the following notation. Let the linear form $\sum_{j=1}^{N} a_{ij}\lambda_j$ be denoted by A_i. Then (4) becomes

$$\left(\sum_{1}^{N} A_i t_i\right)^{(K)} = \left(\sum_{1}^{N} \lambda'_i t_i\right)^{(K)}. \tag{6}$$

Thus in a symbolic sense A_i's and λ_i's are equivalent. But with regard to the subscripts of the A terms in the expansion of the left member of (6) we use a different convention than for the subscripts of the λ's. We set

---

[1] See the author, loc. cit., p. 19

$$\lambda'_{20} = A_{11}\,,\ \lambda_{020} = A_{22}\,,\ \cdots,$$

$$\lambda'_{110} = A_{12}\,,\ \lambda_{1010} = A_{13}\,,\ \cdots$$

We get

$$S_1(V_2) = \frac{1}{N}\Sigma A_{ii}\,,$$

$$S_2(V_2) = \frac{1}{N^2}\left[3\Sigma A_{ii}^2 + 2\Sigma A_{ii}A_{jj} + 4\Sigma A_{ij}^2 - (\Sigma A_{ii})^2\right],\ i \neq j,$$

the summations, of course, running over all values of i and j from 1 to N. But since

$$\Sigma A_{ii}^2 + 2\Sigma A_{ii}A_{jj} = (\Sigma A_{ii})^2$$

the second relation reduces to

$$S_2(V_2) = \frac{2}{N^2}\left(\Sigma A_{ii}^2 + 2\Sigma A_{ij}^2\right).$$

Similarly

$$S_3(V_2) = \frac{8}{N^3}\left(\Sigma A_{ii}^3 + 3\Sigma A_{ii}A_{ij}^2 + 6\Sigma A_{ij}A_{ik}A_{jk}\right),$$

$$S_4(V_2) = \frac{48}{N^4}\left(\Sigma A_{ii}^4 + 4\Sigma A_{ii}^2 A_{ij}^2 + 4\Sigma A_{ii}A_{jj}A_{ij}^2 + 2\Sigma A_{ij}^4\right.$$

$$\left. + 8\Sigma A_{ii}A_{ij}A_{ik}A_{jk} + 4\Sigma A_{ij}^2 A_{ik}^2 + 8\Sigma A_{ij}A_{ik}A_{jl}A_{kl}\right), \tag{7}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S_1(V_3) = 0,$$

$$S_2(V_3) = \frac{3}{N^2}\left(5\Sigma A_{ii}^3 + 6\Sigma A_{ii}A_{jj}A_{ij} + 4\Sigma A_{ij}^3\right),$$

$$S_3(V_3) = 0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S_1(V_4) = \frac{3}{N}\Sigma A_{ii}^2 ,$$

$$S_2(V_4) = \frac{48}{N^2}\left(2\Sigma A_{ii}^4 + 3\Sigma A_{ii} A_{jj} A_{ij}^2 + \Sigma A_{ij}^2\right).$$

To illustrate the use of these formulas and to give some results in a case of practical interest, let us suppose that the set of N observations composing a sample may be assigned an order in which only consecutive observations are correlated and in a constant degree. Thus our observations might be prices or indices taken at the ends of consecutive time intervals. We suppose, then, that

$$\lambda_{110} = \lambda_{0110} = \lambda_{00110} = \cdots = r\lambda_{20} ,$$

$$\lambda_{101} = \lambda_{1001} = \lambda_{0101} = \cdots = 0$$

The first step in the calculation is to obtain the values of the various A's which enter into the formulas (7). A_{11} is found from A_1^2, A_{12} from $A_1 A_2$ and so on. We get

$$
\begin{aligned}
& A_{11} = A_{N,N} = \left(1 - \frac{1}{N} - \frac{2r}{N^2}\right)\lambda_{20} , \\
& A_{22} = A_{33} = \cdots = A_{N-1,N-1} = \left(1 - \frac{1+2r}{N} - \frac{2r}{N^2}\right)\lambda_{20} , \\
& A_{12} = A_{N-1,N} = \left(r - \frac{1+r}{N} - \frac{2r}{N^2}\right)\lambda_{20} , \\
(8)\quad & A_{23} = A_{34} = \cdots = A_{N-2,N-1} = \left(r - \frac{1+2r}{N} - \frac{2r}{N^2}\right)\lambda_{20} , \\
& A_{13} = A_{14} = \cdots = A_{1,N-1} , \\
& \quad = A_{2,N} = A_{3,N} = \cdots = A_{N-2,N} = \left(-\frac{1+r}{N} - \frac{2r}{N^2}\right)\lambda_{20} , \\
& A_{1,N} = \left(-\frac{1}{N} - \frac{2r}{N^2}\right)\lambda_{20} , \\
& A_{ij} = \left(-\frac{1+2r}{N} - \frac{2r}{N^2}\right)\lambda_{20} \qquad \begin{cases} 1 < i < N-1 \\ 1 < j < N-1 \\ |i-j| > 1 \end{cases}
\end{aligned}
$$

Then, on substitution in (7), we have finally

$$S_1(V_2) = \left(1-\frac{1}{N}\right)\left(1-\frac{2r}{N}\right)\lambda_{20},$$

$$S_2(V_2) = \frac{2}{N}\left[\left(1-\frac{1}{N}\right)\left(1-\frac{4r}{N}\right)+2r^2\left(1-\frac{3}{N}+\frac{2}{N^2}+\frac{2}{N^3}\right)\right]\lambda_{20}^2$$

These two results are given by Dr. Rhodes, loc cit, though there is a slight misprint in the second one as given there The remainder of the results given here are believed to be new,

$$S_3(V_2) = \frac{8}{N^2}\left[\left(1-\frac{1}{N}\right)\left(1-\frac{6r}{N}\right)+6r^2\left(1-\frac{3}{N}+\frac{2}{N^2}+\frac{2}{N^3}\right)-\frac{4r^3}{N}\left(2-\frac{3}{N}-\frac{3}{N^2}-\frac{2}{N^3}\right)\right]\lambda_{20}^3,$$

$$S_1(V_3) = 0,$$

$$S_2(V_3) = \frac{6}{N}\left[\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\left(1-\frac{6r}{N}\right)-\frac{6r^2}{N}\left(1-\frac{5}{N}+\frac{12}{N^3}\right)+\frac{2r^3}{N^2}\left(1-\frac{7}{N}+\frac{14}{N^2}\right.\right.$$

$$\left.\left.+\frac{2}{N^3}-\frac{24}{N^4}-\frac{40}{N^5}\right)\right]\lambda_{20}^4,$$

$$S_3(V_3) = 0,$$

$$S_1(V_4) = 3\left[\left(1-\frac{1}{N}\right)^2\left(1-\frac{4r}{N}\right)+\frac{4r^2}{N^2}\left(1-\frac{3}{N^2}\right)\right]\lambda_{20}^2,$$

$$S_2(V_4) = \frac{24}{N}\left[\left(1-\frac{1}{N}\right)\left(4-\frac{9}{N}+\frac{6}{N^2}\right)\left(1-\frac{8r}{N}\right)+6r^2\left(1-\frac{5}{N}+\frac{25}{N^2}-\frac{29}{N^3}-\frac{44}{N^4}+\frac{68}{N^5}\right)\right.$$

$$-\frac{8r^3}{N}\left(4-\frac{19}{N}+\frac{33}{N^2}+\frac{30}{N^3}-\frac{54}{N^4}-\frac{108}{N^5}\right)+2r^4\left(1-\frac{9}{N}+\frac{44}{N^2}-\frac{64}{N^3}\right.$$

$$\left.\left.-\frac{114}{N^4}+\frac{192}{N^5}+\frac{360}{N^6}+\frac{288}{N^7}\right)\right]\lambda_{20}^4.$$

It should be observed that the expressions for $S_1(V_n)$ for $N < 3$ and for $S_k(V_n)$, $k \geq 2$ for $N < 5$ are in general not valid, since it can be seen by reference to (8) that all the types of A's used in the formulas (7) do not exist for values of N so small. But for these small values of N, the values of the characteristics for which expressions are given above can be readily computed directly.

Stanford University

THE RELATION BETWEEN THE MEANS AND VARIANCES, MEANS SQUARED AND VARIANCES IN SAMPLES FROM COMBINATIONS OF NORMAL POPULATIONS

By

G A BAKER

The distributions of the means and variances of samples from the combinations of normal populations have been discussed in a previous paper [1] It is known that if the sampled population is not normal the means and variances of samples are not independent

The present discussion aims to give some idea of the relation between the means and the variances, means squared and variances of samples from a population that is the combination of normal populations To this end the case of samples of two from such populations is rather completely investigated. Also empirical random sampling results for two special populations are presented

Suppose that a population is represented by

$$(1) \qquad f(x) = \frac{1}{1+k}\left[\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} + \frac{k}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}\right].$$

[1]"Random Sampling from Non-Homogeneous Populations," Metron, Vol VIII, No 3 (1930), pp 1-21

If a method used by Karl Pearson[1] is followed, the probability of

$$x_1 \text{ in } dx_1 \text{ is } f(x_1)dx_1 ,$$
$$x_2 \text{ in } dx_2 \text{ is } f(x_2)dx_2$$

and the probability of the concurrence of these two events is

$$(2) \qquad f(x_1)\,f(x_2)\,dx_1\,dx_2$$

which may be written

$$(3) \qquad \frac{1}{(1+k)^2 2\pi}\left[e^{-\frac{1}{2}\left[x_1^2+x_2^2\right]}+\frac{k^2}{\sigma^2}\,e^{-\frac{1}{2\sigma^2}\left[(x_1-m)^2+(x_2-m)^2\right]} + \frac{k}{\sigma}\left\{e^{-\frac{1}{2}\left[x_1^2+\frac{(x_2-m)^2}{\sigma^2}\right]}+e^{-\frac{1}{2}\left[x_2^2+\frac{(x_1-m)^2}{\sigma^2}\right]}\right\}\right]dx_1dx_2$$

Now

$$x = \frac{1}{2}(x_1+x_2)$$

$$\Sigma^2 = \frac{1}{2}\left[(x_1-x)^2+(x_2-x)^2\right]$$

Whence

$$(4) \qquad \begin{cases} x_1 = -\Sigma + x \\ x_2 = \Sigma + x \end{cases}$$

Also $dx_1\,dx_2$ may be replaced[2] by

[1] Appendix to Papers by "Student" and R A Fisher, Biometrika, Vol XIX (1925), p 522

[2] R A Fisher "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," Biometrika, Vol X (1915), p 507

$$c\, d\Sigma\, dx \tag{5}$$

In virtue of (5), (4) and (3), (6) is obtained

$$\frac{1}{(1+k)^2 2\pi}\left[e^{-\frac{1}{2}[2x^2+2\Sigma^2]} + \frac{k^2}{\sigma^2} e^{-\frac{1}{2\sigma^2}[2\Sigma^2+2(x-m)^2]}\right.$$

$$+\frac{k}{\sigma}\left\{e^{-\frac{1}{2}\left[(-\Sigma+x)^2+\frac{(\Sigma+x-m)^2}{\sigma^2}\right]}\right. \tag{6}$$

$$\left.\left.+e^{-\frac{1}{2}\left[(\Sigma+x)^2+\frac{(-\Sigma+x-m)^2}{\sigma^2}\right]}\right\}\right].$$

This is the correlation surface for the means and standard deviations of samples of two drawn from (1). To get the correlation surface of the means and variances write

$$\Sigma^2 = u$$

$$d\Sigma = \frac{du}{2\sqrt{u}}$$

Then

$$F(x,u) = \frac{1}{(1+k)^2 2\pi}\left[\frac{e^{-\frac{1}{2}[2x^2+2u]}}{2\sqrt{u}} + \frac{k^2}{2\sqrt{u}\,\sigma^2} e^{-\frac{1}{2\sigma^2}[2u+2(x-m)^2]}\right.$$

$$+\frac{k}{\sigma}\left\{\frac{e^{-\frac{1}{2}\left[(\sqrt{u}+x)^2+\frac{(-\sqrt{u}+x-m)^2}{\sigma^2}\right]}}{2\sqrt{u}}\right. \tag{7}$$

$$\left.\left.+\frac{e^{-\frac{1}{2}\left[(-\sqrt{u}+x)^2+\frac{(\sqrt{u}+x-m)^2}{\sigma^2}\right]}}{2\sqrt{u}}\right\}\right]$$

is the desired surface.

The locus of mean u's for given x's is

$$u=\frac{e^{-x^2}+k^2\sigma e^{-\frac{(x-m)^2}{\sigma^2}}+\frac{4\sqrt{2}k}{(\sigma^2+1)^{\frac{3}{2}}}e^{-\frac{2\left(x-\frac{m}{2}\right)^2}{\sigma^2+1}}\left[\sigma^2+\frac{\left((\sigma^2-1)x+m\right)^2}{\sigma^2+1}\right]}{e^{-x^2}+\frac{k^2}{\sigma}e^{-\frac{(x-m)^2}{\sigma^2}}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}e^{-\frac{2\left(x-\frac{m}{2}\right)^2}{\sigma^2+1}}}. \tag{8}$$

The locus of the mean x's for given u's is

$$x=\frac{\frac{mk^2}{\sigma}e^{-\frac{u}{\sigma^2}}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}\left[\left\{(\sigma^2-1)\sqrt{u}+m\right\}e^{-\frac{2\left(\sqrt{u}-\frac{m}{2}\right)^2}{\sigma^2+1}}-\left\{(\sigma^2-1)\sqrt{u}-m\right\}e^{-\frac{2\left(\sqrt{u}+\frac{m}{2}\right)^2}{\sigma^2+1}}\right]}{e^{-u}+\frac{k^2}{\sigma}e^{-\frac{u}{\sigma^2}}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}\left\{e^{-\frac{2\left(\sqrt{u}-\frac{m}{2}\right)^2}{\sigma^2+1}}+e^{-\frac{2\left(-\sqrt{u}-\frac{m}{2}\right)^2}{\sigma^2+1}}\right\}}. \tag{9}$$

The correlation surface for the means squared (= z) and variances is

$$\begin{aligned}\psi(u,z)=\frac{1}{(1+k)^2 2\pi}\Bigg[&\frac{e^{-\frac{1}{2}[2z+2u]}}{4\sqrt{u}\sqrt{z}}+\frac{k^2e^{-\frac{1}{2\sigma^2}\left[2u+2(\sqrt{z}-m)^2\right]}}{\sigma^2 4\sqrt{u}\sqrt{z}}\\ &+\frac{k}{\sigma}\Bigg\{\frac{e^{-\frac{1}{2}\left[(\sqrt{u}+\sqrt{z})^2+\frac{(-\sqrt{u}+\sqrt{z}-m)^2}{\sigma^2}\right]}}{4\sqrt{u}\sqrt{z}}\\ &+\frac{e^{-\frac{1}{2}\left[(-\sqrt{u}+\sqrt{z})^2+\frac{(\sqrt{u}+\sqrt{z}-m)^2}{\sigma^2}\right]}}{4\sqrt{u}\sqrt{z}}\Bigg\}\Bigg].\end{aligned} \tag{10}$$

The locus of the mean u's for given z's is

$$(11)\quad u=\frac{e^{-z}+k^{2}\sigma e^{-\frac{(\sqrt{z}-m)^{2}}{\sigma^{2}}}+\frac{4\sqrt{2}\,k}{(\sigma^{2}+1)^{\frac{3}{2}}}e^{-\frac{2(\sqrt{z}-\frac{m}{2})^{2}}{\sigma^{2}+1}}\left[\sigma^{2}+\frac{\{(\sigma^{2}-1)\sqrt{z}+m\}^{2}}{\sigma^{2}+1}\right]}{e^{-z}+\frac{k^{2}}{\sigma}e^{-\frac{(\sqrt{z}-m)^{2}}{\sigma^{2}}}+\frac{2\sqrt{2}\,k}{\sqrt{\sigma^{2}+1}}e^{-\frac{2(\sqrt{z}-\frac{m}{2})^{2}}{\sigma^{2}+1}}}$$

The locus of the mean z's for given u's is

$$z=\frac{1}{e^{-u}+\frac{k^{2}}{\sigma}e^{-\frac{u}{\sigma^{2}}}+\frac{2\sqrt{2}\,k}{\sqrt{\sigma^{2}+1}}\left\{e^{-\frac{2(\sqrt{u}-\frac{m}{2})^{2}}{\sigma^{2}-1}}+e^{-\frac{2(\sqrt{u}+\frac{m}{2})^{2}}{\sigma^{2}+1}}\right\}}$$

(12) multiplied by

$$\Bigg[e^{-u}+\frac{k^{2}}{\sigma}(\sigma^{2}+m^{2})e^{-\frac{u}{\sigma^{2}}}+\frac{4\sqrt{2}\,k}{(\sigma^{2}+1)^{\frac{3}{2}}}\Bigg\{\left(\sigma^{2}+\frac{[(\sigma^{2}-1)\sqrt{u}-m]^{2}}{\sigma^{2}+1}\right)e^{-\frac{2(\sqrt{u}-\frac{m}{2})^{2}}{\sigma^{2}+1}}+\left(\sigma^{2}+\frac{[(\sigma^{2}-1)\sqrt{u}+m]^{2}}{\sigma^{2}+1}\right)e^{-\frac{2(\sqrt{u}+\frac{m}{2})^{2}}{\sigma^{2}+1}}\Bigg\}\Bigg]$$

By expanding the denominators of (8), (9), (11), and (12) by the multinomial theorem, it can be shown that each of these loci is essentially parabolic, $\sigma^{2}-1 \neq 0$ They are subject to an exponential influence at the beginning of the range of the independent variable, which influence rapidly diminishes as the independent variable takes on higher values

The probability relations in general between means and variances, means squared and variances will be expected to approximate those for the case of samples of two, because of the fol-

lowing considerations Suppose that n (the number in the sample) is large[1] When a large proportion of the sample comes from the first component, the first term of (7) with 2 in the numerator of the exponent replaced by n and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances Similarly, when a large proportion of the sample comes from the second component, the second term of (7) with 2 in the numerator of the exponent replaced by n and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances When about equal proportions of the sample come from each component, the last term of (7) with $\frac{2}{2}$ in the numerator of each exponent replaced by $\frac{n}{2}$ and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances. Or, all together, (7) with the mentioned changes in the exponents of the terms, with proper weighting of the terms, and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ is a proportionate approximation to the distribution of the means and variances of samples drawn from a population represented by (1) Further, increasing n will not influence relations (8), (9), (11), and (12) as approximations for the general case except the exponential term, if it is assumed that the denominators are expanded and then multiplied by the numerators, for $\frac{1}{n}$ occurs to the same power in the numerators and denominators

---

[1]Note The effect of k and of the binomial coefficients is roughly as follows If the $n+1$ terms denoting s from the first component of (1) and $n-s$ from the second component are divided into thirds, then, if l_1, l_2, l_3 are the exponents of k in the middle terms,

$$l_1 = \frac{n-2}{6}, \quad l_2 = \frac{3n}{6}, \quad l_3 = \frac{5n+2}{6}$$

or approximately, since n is large and since only a proportionate expression is desired $l_1' = 0$, $l_2' = \frac{n}{3}$, $l_3 = \frac{2n}{3}$ or the exponents of k of the middle terms of the three sections above are $\frac{n}{3}$ times the exponents of k in (7). The effect of increasing n because of the binomial coefficients is to weight the middle section of the possible surfaces to a much greater extent than the extreme sections, so that with n very large the last term of (7) with 2 replaced by n becomes an approximation to the desired surface

From (8), (9), (11), and (12) it is clear that the parameters of the sampled population have great influence on the regression relations considered It should be borne in mind in this connection that many flattened and skewed, as well as bimodal, distributions can be adequately represented by combinations of normal populations Also, results (8), (9), (11), and (12) can be extended to the sums and differences of any number of normal curves, subject to the condition that the resultant is always positive

In 1925, Dr Neyman[1] gave the correlation coefficient between the deviations of the means of samples from the mean of the sampled population and the variances of these samples for samples of n drawn at random from an infinite uni-variate population in terms of the betas of the sampled population as

$$(13) \qquad \rho' = \frac{\sqrt{n-1}\,\sqrt{\beta_1}}{\sqrt{(n-1)\beta_2 - n + 3}} .$$

Similarly, the correlation coefficient between the deviations squared of the means of samples from the mean of the sampled population and the variances is

$$(14) \qquad R' = \frac{\sqrt{n-1}\,(\beta_2 - 3)}{\sqrt{(\beta_2 + 2n - 3)\left[(n-1)\beta_2 - n + 3\right]}}$$

Under certain very special conditions the statement of ρ' and R' may give an adequate idea of the regression relation between the means and variances, means squared and variances of samples from a population represented by (1) In general the mere statement of these coefficients will not give any useful

[1]J Splawa-Neyman "Contributions to the Theory of Small Samples Drawn from a Finite Population," Biometrika, Vol XVII (1925), pp 472-479

notion of the actual probability relations This is true because (a) the regression relations between means and variances, means squared and variances of samples from a population represented by (1) are essentially parabolic, as shown for samples of two and as seems probable for larger samples, (b) the frequency arrays may vary markedly in dispersion, in skewness, and in other characteristics

To illustrate these remarks, samples of four were drawn from two special populations by throwing dice

Suppose that a population is represented by

$$(15) \quad f(x)=\frac{1}{1+k}\left[\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x+m_1)^2}+\frac{k}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-m_2)^2}{\sigma^2}}\right].$$

The first four moments of $f(x)$ about its mean are

$$u_0 = 1,$$

$$u_1 = \frac{(-m_1+km_2)}{1+k} = 0,$$

$$u_2 = \frac{[1+m_1^2+k(\sigma^2+m_2^2)]}{1+k},$$

$$u_3 = \frac{-3m_1-m_1^3+k(3m_2\sigma^2+m_2^3)}{1+k},$$

$$u_4 = \frac{3+6m_1^2+m_1^4+k(3\sigma^4+6m_2^2\sigma^2+m_2^4)}{1+k}.$$

CHART A

Population I, from Which the 1038 Samples of Four of Tables I and II were drawn

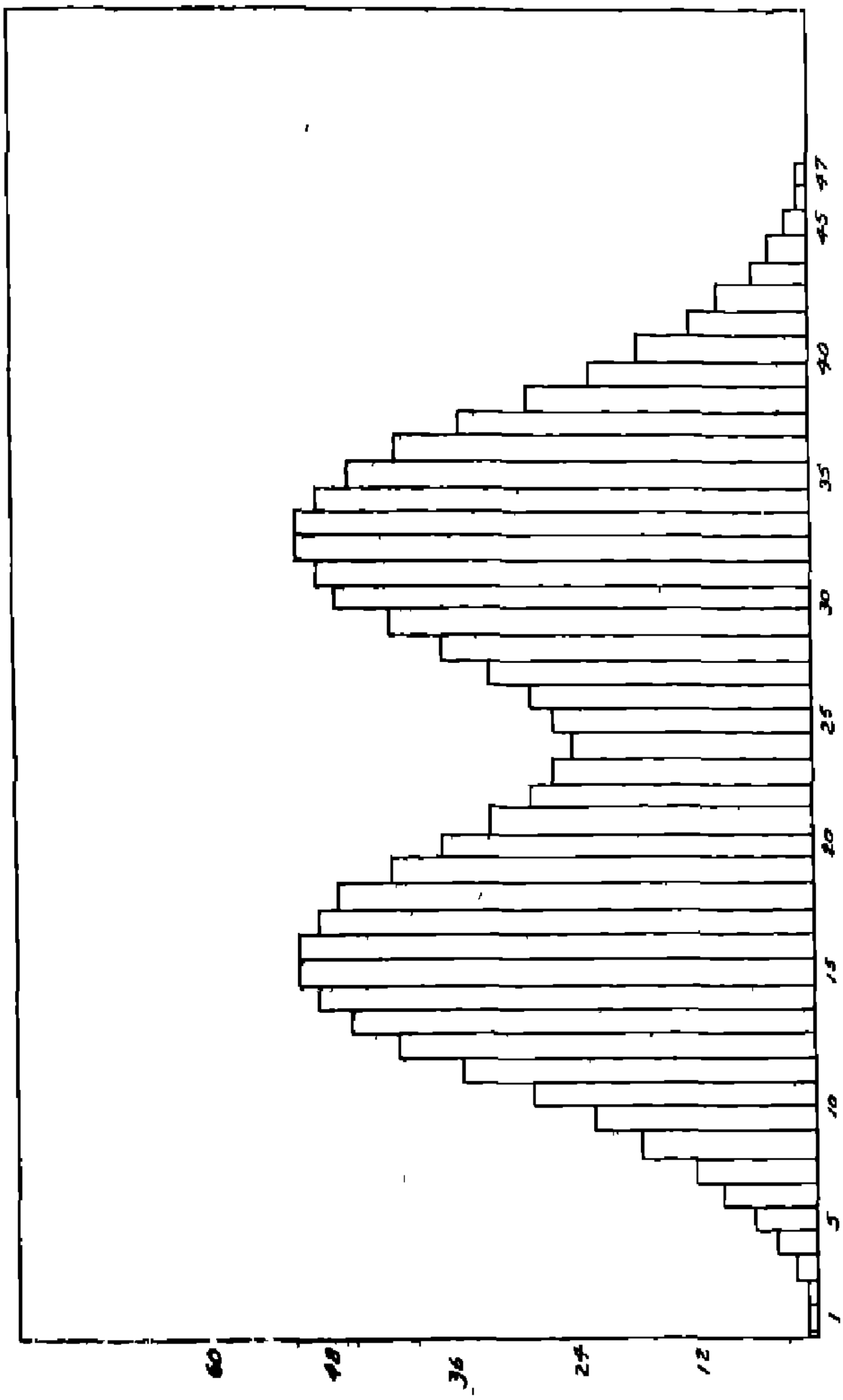


TABLE I

Correlation Table Showing the Relation Between the Means and Variances of Samples of Four from Population I

Means of Samples	Variances of Samples: 0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'tls
15 to 17			1															1
13 to 15	1	1																2
11 to 13	5	1	1	2														9
9 to 11	13	8	0	4		2	2											29
7 to 9	11	15	10	5	7	2	6	7			1		1					65
5 to 7	4	13	16	22	10	7	11	7	1	2								93
3 to 5	12	13	12	12	25	12	5	10	4	2	2	3	1	1			1	115
1 to 3	4	8	18	18	20	16	32	14	8	5	3	3	2	1		1		153
− 1 to 1	5	4	7	34	21	24	27	29	8	10	7	5	1		2			184
− 3 to − 1	5	12	14	14	16	18	21	13	7	7	3	2		1	1	1		135
− 5 to − 3	2	17	8	18	12	18	17	10	6	1	5	1	1	1	1			118
− 7 to − 5	6	6	12	3	14	9	3	6	4	1			1					65
− 9 to − 7	11	6	9	3	9	2	1	2	2									45
−11 to − 9	3	6	3				1											13
−13 to −11	4	1	1		1													7
−15 to −13	1	1	1		1													4
−17 to −13																		
Totals	87	112	113	135	136	110	126	98	40	28	21	14	7	4	4	2	1	1038

Whence

$$(16)\quad \beta_1 = \frac{(1+k)[-3m_1 - m_1^3 + k(3m_2\sigma^2 + m_2^3)]^2}{[1+m_1^2+k(\sigma^2+m_2^2)]^3},$$

$$(17)\quad \beta_2 = \frac{(1+k)[3+6m_1^2+m_1^4+k(3\sigma^4+6m_2^2\sigma^2+m_2^4)]}{[1+m_1^2+k(\sigma^2+m_2^2)]^2}.$$

Thus, for any special population of the form (15), β' and ρ' can be easily calculated

Samples of four were drawn from a population approximately represented by

$$(18)\qquad f_1(x) = 648\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x+1.7)^2} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-1.7)^2}\right].$$

The actual sampled population is shown in Chart A and is hereinafter called Population I

Table I shows the distribution of 1038 samples of four drawn from Population I with respect to the observed values of the means and the variances. The arrays for constant values of the variances are at first distinctly bimodal, gradually becoming unimodal Chart I shows the means of arrays of Table I with the regression lines as calculated without correction for groupings It is apparent that the locus of the mean variances for a given value of the means diverges a great deal from a straight line This regression relation looks as though it was a normal curve,

which is what would be expected from (8) with $\sigma^2 - 1 = 0$ The theoretical and actual correlation coefficients for this and three subsequent tables are compared in Table V and the constants of the marginal distributions of Tables I to IV are presented in Table VI

If the deviations of the means of the samples of Table I from the mean of Population I are squared, Table II results, Chart II shows the means of arrays and regression lines of Table II. The regression lines are very poor fits to the means of the arrays which are, apparently, exponential loci

Table III shows the distribution of 1058 samples of four drawn from a population approximately represented by

$$(19) \quad f_2(x) = 972\left[\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x+8)^2} + \frac{1}{3\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-24)^2}\right]$$

with respect to the observed values of the means and variances of the samples, The actual sampled population is presented in Chart B and is hereinafter called Population II Chart III shows the means of arrays and regression lines of Table III This chart resembles Chart I in that the locus of the mean variances for given values of the means is so obviously non-linear Also, a glance at Table III is sufficient to see that the arrays vary markedly in skewness

Table IV shows the relation between the means squared and variances of samples of four from Population II Chart IV shows the means of arrays and regression lines for Table IV In this case the regression relations seem to be fairly near linear, and the frequency distributions of the arrays do not change strikingly

CHART I

The Means of Arrays and Regression Lines of the Means and Variances of Samples from Population I

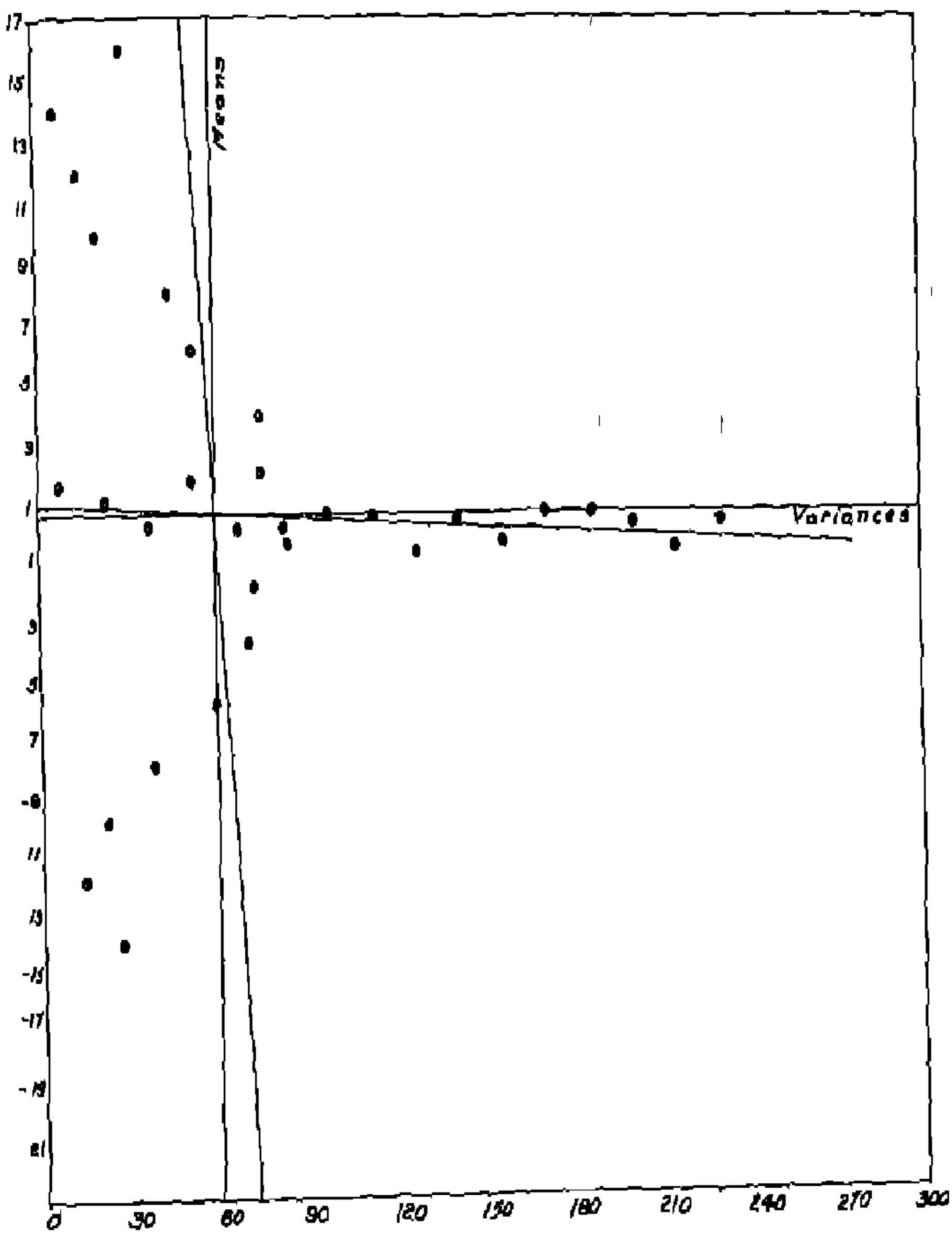


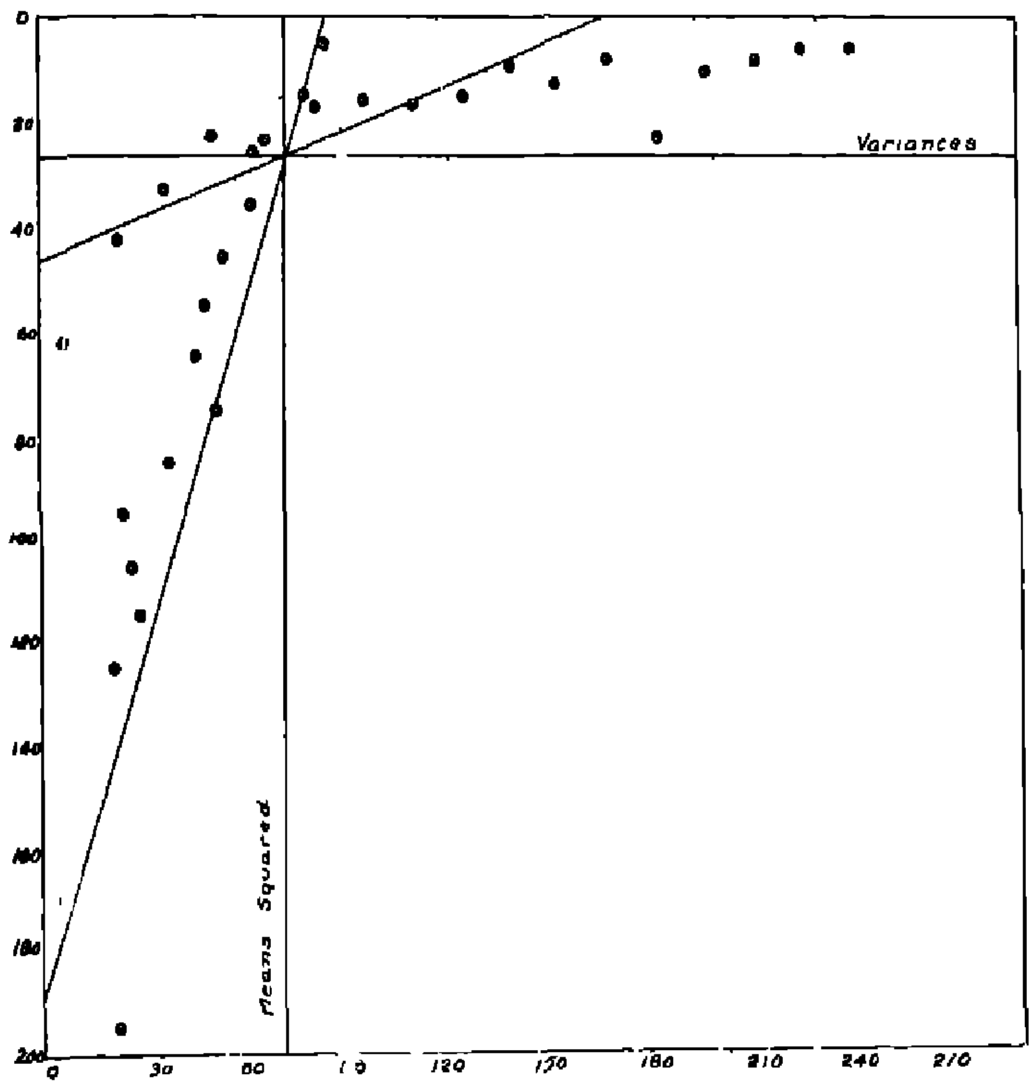
TABLE II

Correlation Table Showing the Relation Between the Means Squared and Variances of Samples of Four from Population I

Means Squared of Samples	Variances of Samples																	
	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
0 to 10	13	25	42	64	59	58	83	56	25	22	12	10	3	2	3	2	1	480
10 to 20	7	21	10	18	24	19	14	15	6	3	8	4	2	2				154
20 to 30	5	14	13	21	16	15	11	9	2									106
30 to 40	4	9	13	12	17	9	6	5	2	3								80
40 to 50	7	7	11	6	5	4	2	5	3				1					51
50 to 60	7	5	6	2	5			4	2									31
60 to 70	10	6	6	4	6	2	5	2										41
70 to 80	6	6	3	2	1	1	2	2			1		1					25
80 to 90	5	3	1		2	2	1											14
90 to 100	5	3	2	2														12
100 to 110	4	6	1	2			1											14
110 to 120	3	1	1				1											6
120 to 130			1															2
130 to 140	5	2	2	1	1													11
140 to 150	2	1																3
150 to 160																		
160 to 170	1																	1
170 to 180		2	1															3
180 to 190	2																	2
190 to 200																		
220 to 230				1														1
250 to 260		1																1
Totals	87	112	113	135	136	110	126	98	40	28	21	14	7	4	4	2	1	1038

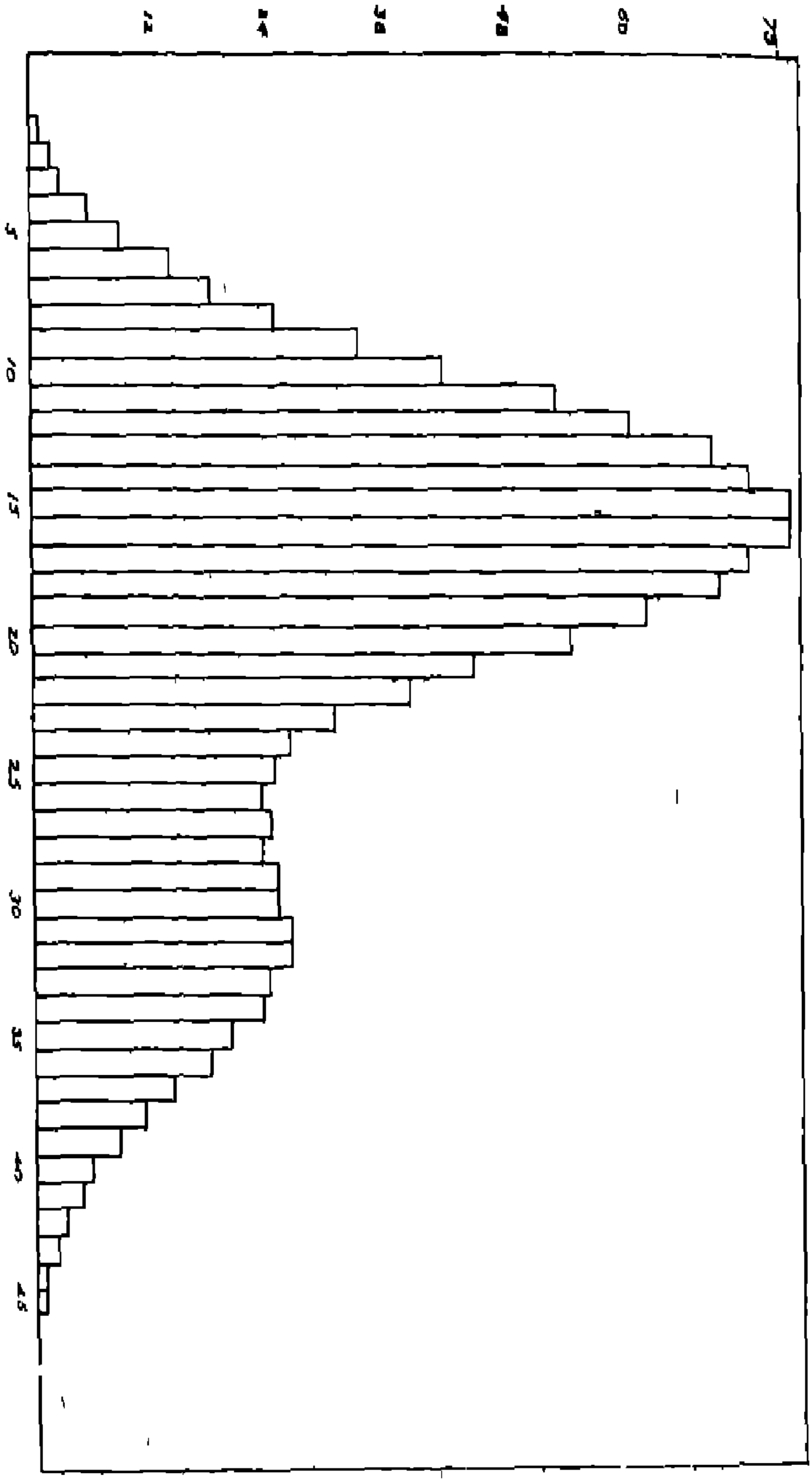
CHART II

The Means of Arrays and Regression Lines of the Means Squared and Variances of Samples from Population I



NOTE The last thirteen class intervals of the means squared are grouped into one group

CHART B



Population II, from Which the 1058 Samples of Four of Tables III and IV Were Drawn

TABLE III

Correlation Table Showing the Relation Between the Means and Variances of Samples of Four from Population II

Means of Samples	Variances of Samples: 0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
15 to 17	1		1															2
13 to 15		2	1															3
11 to 13			2				1	1										4
9 to 11	2	2	3		5	1	0	4					1					18
7 to 9	3	3	10	4	6	7	2	4	5	4		1						49
5 to 7	5	3	13	17	6	10	4	6	10	2	2	4			1			83
3 to 5	[illegible]	11	12	18	24	12	18	9	2	6	3	2						121
1 to 3	9	13	21	22	19	27	7	7	5	7	6		1		1	1		146
− 1 to 1	24	26	29	24	16	13	15	8	8	1	4	3	1		1			173
− 3 to − 1	45	36	26	22	14	14	11	4	5	1	1	1	1			1		182
− 5 to − 3	59	36	21	15	10	8			1	2								152
− 7 to − 5	36	29	7	7	2	1	1	1			1							85
− 9 to − 7	14	9	10		1													34
−11 to − 9	3	1																4
−13 to −11	1	1																2
−15 to −13																		
−17 to −15																		
Totals	206	172	156	129	103	93	59	44	36	23	17	11	4		3	2		1058

CHART III

The Means of Arrays and Regression Lines of the Means and Variances of Samples from Population II

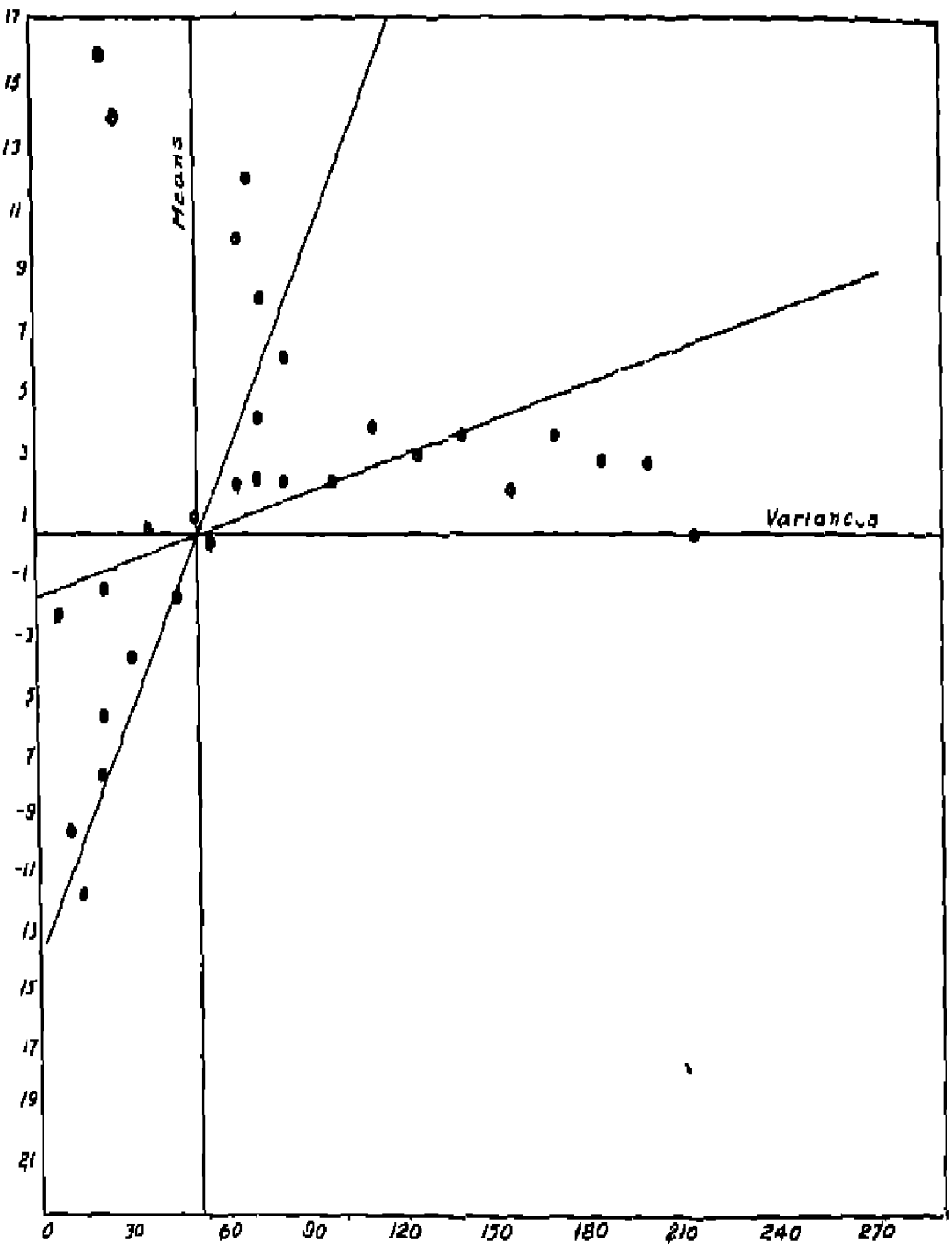


TABLE IV

Correlation Table Showing the Relation Between the Means Squared and Variances of Samples of Four from Population II

Means Squared of Samples	Variances of Samples: 0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
0 to 15	117	101	93	89	68	67	46	24	19	13	12	5	4		2	2		661
15 to 30	36	35	24	21	17	12	5	7	5	5	3	2						172
30 to 45	26	16	7	15	5	6	5	3	7		2	3			1			96
45 to 60	10	9	14	2	6	2	1	2	1	2	0	1						50
60 to 75	5	4	9	2	1	3	1	3	3	2								33
75 to 90	6	3	3		2	2		3	1	1								21
90 to 105	3		2		2	1		1					1					10
105 to 120	1	1			2													4
120 to 135	1																	2
135 to 150		1					1											2
150 to 165			2															2
165 to 180			1															1
180 to 195		1																1
195 to 210																		
210 to 225		1																1
225 to 240	1		1															2
240 to 255																		
Totals	206	172	156	129	103	93	59	44	36	23	17	11	4		3	2		1058

CHART IV

The Means of Arrays and Regression Lines of the Means Squared and Variances of Samples from Population II

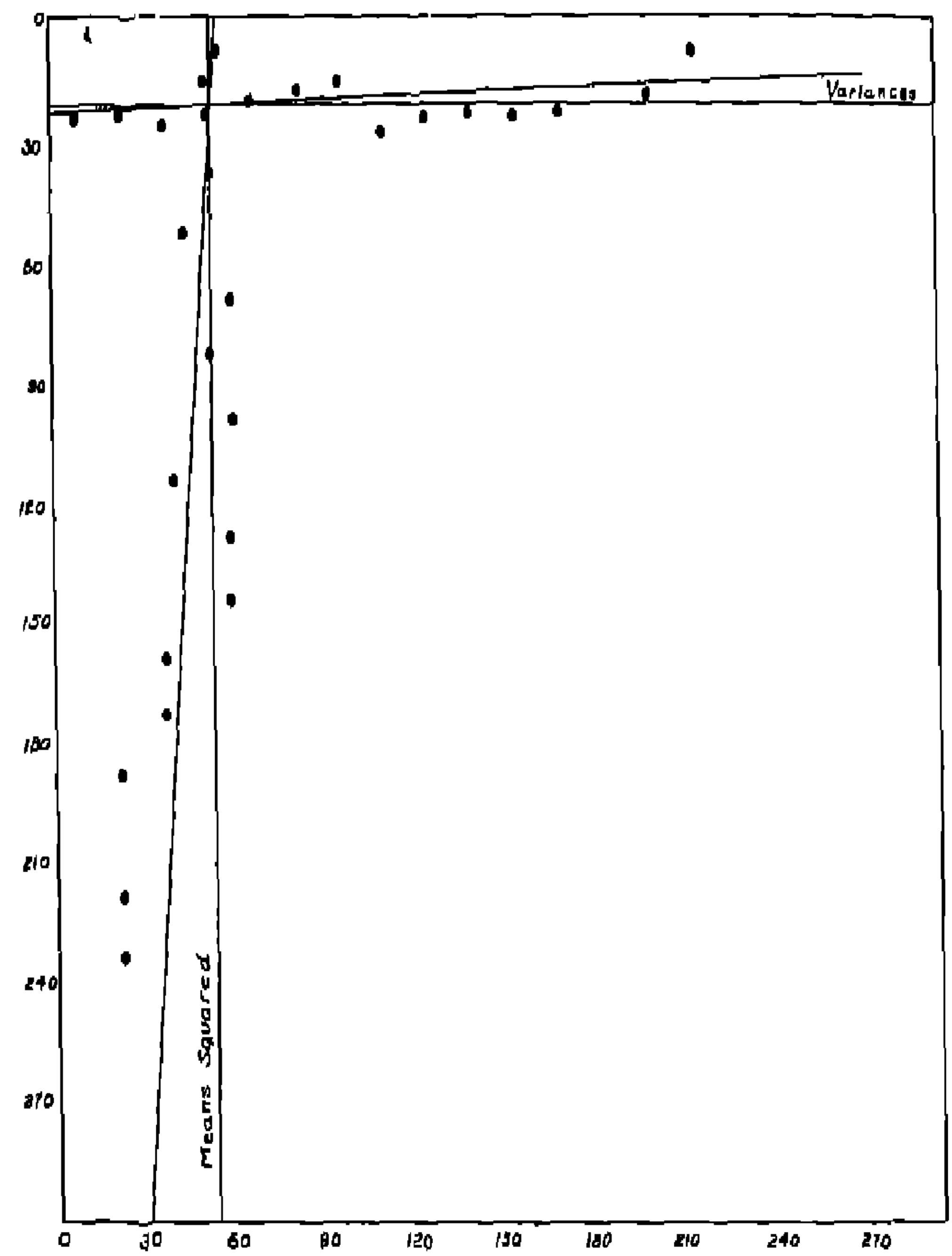


TABLE V

Correlation Coefficients of Tables I-IV

	Correlation-Coefficient	
Number of Table	Theoretical	Actual[1]
I	00	-05
II	-34	-37
III	40	37
IV	-07	-05

TABLE VI

Constants of the Marginal Distributions of Tables I-IV in Terms of Class Intervals

Marginal Distribution	Mean	Standard Deviation
Means of Samples from Population I	252[2]	2 467
Variances of Samples from Population I	4 890[3]	2 900
Means Squared of Samples from Population I	3 591[3]	3 203
Means of Samples from Population II	07	2 237
Variances of Samples from Population II	3 570[3]	2 854
Means Squared of Samples from Population II	1 408[3]	1 744

[1] Calculated without corrections for grouping

[2] Is so far from zero because of the groupings employed Many means were exactly odd integers These were all put forward into higher classes making the calculated mean too large

[3] Origin taken at the beginning of the range

From the results for the case of samples of two and from the results of empirical sampling, it seems clear that the simplest regression relation that is generally applicable to the means and variances, means squared and variances, of samples from populations which are the combinations of normal populations is parabolic. For small samples and for certain values of the parameters of the sampled population the regression relations may involve exponential terms that are quite important. As the size of the samples increases, it is expected that this exponential term will decrease in influence. It seems plausible that even with large samples the regression relation of means and variances, means squared and variances will remain essentially parabolic. It is not expected that the determination of a good approximation to the regression relations will serve to give an adequate notion of the probability relations of the means and variances, means squared and variances of samples from a population represented by (1), because the arrays may vary in number of modes, in skewness, in dispersion, and in other characteristics. For instance, surface (7) may be trimodal so that arrays may be bimodal or unimodal, and in such a case the arrays must vary markedly. Surfaces (7) and (10) with 2 replaced by n and with the terms suitably weighted are valuable approximations to the probability relations of the means and variances, means squared and variances of samples drawn from a population represented by (1).

G A Baker

A TABLE TO FACILITATE THE FITTING OF CERTAIN LOGISTIC CURVES

By

JOSHUA L BAILEY, JR

The most useful generalization of the logistic curve is that having the form

$$(1) \qquad y = \frac{k}{1+e^{a+bx+cx^2+gx^3}}$$

In practice it will seldom be found necessary to use higher powers of x This equation may also be written

$$(2) \qquad Y = a + bx + cx^2 + gx^3$$

in which $Y \equiv \log \frac{k-y}{y}$

If we can evaluate the constant k with reasonable accuracy, the value of Y corresponding to each observed value of y can be computed, and then the values of the coefficients a, b, c, and g, in equation (1) may be obtained by fitting equation (2) as a generalized parabola by the method of least squares

The normal equations necessary to make this fit will be found to be

$$a\Sigma x^0 + b\Sigma x + c\Sigma x^2 + g\Sigma x^3 = \Sigma Y$$
$$a\Sigma x + b\Sigma x^2 + c\Sigma x^3 + g\Sigma x^4 = \Sigma xY$$
$$a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 + g\Sigma x^5 = \Sigma x^2 Y$$
$$a\Sigma x^3 + b\Sigma x^4 + c\Sigma x^5 + g\Sigma x^6 = \Sigma x^3 Y.$$

In the special case where the observations have been made at regular intervals (that is, where the successive values of x are in arithmetic progression) the solution of these normal equations may be greatly simplified. We may then select an arbitrary origin in the middle of the range of observations, so that for every positive value of x there will be a corresponding negative value of equal absolute magnitude. Thus the sums of the odd powers of x will all be zero.

If the number of observations be odd, the middle one will, of course, be chosen for the origin, and the unit of the scale will be the interval between successive values of x. If the number of observations be even, the origin will be midway between the middle pair of observations, and it will be found more convenient to take half the interval as scale unit. In the former case, x will take all integral values between $+n$ and $-n$, while in the latter case x may take only the odd integral values.

If we set the sums of the odd powers of x in the normal equations equal to zero, and solve them simultaneously, we derive the following formulae for the literal coefficients.

$$A = \frac{\Sigma Y\,\Sigma X^4 - \Sigma X^2 Y\,\Sigma X^2}{\Sigma X^4 \cdot \Sigma X^0 - (\Sigma X^2)^2}, \quad C = \frac{\Sigma X^2 Y\,\Sigma X^0 - \Sigma Y\,\Sigma X^2}{\Sigma X^4\,\Sigma X^0 - (\Sigma X^2)^2},$$

$$B = \frac{\Sigma XY\,\Sigma X^6 - \Sigma X^3 Y\,\Sigma X^4}{\Sigma X^6\,\Sigma X^2 - (\Sigma X^4)^2}, \quad G = \frac{\Sigma X^3 Y\,\Sigma X^2 - \Sigma XY\,\Sigma X^4}{\Sigma X^6\,\Sigma X^2 - (\Sigma X^4)^2}$$

The use of capital letters indicates that the equation has been referred to the arbitrary origin.

In these formulae the factors involving Y must be computed from the observations, but those in which X alone occurs may be tabulated for all convenient values of n. Since Y does not occur in the denominators at all, these may be tabulated in the same way.

TABLE TO BE USED WHEN THE NUMBER OF OBSERVATIONS IS ODD

n	Σx^0	Σx^2	Σx^4	Σx^6	$\Sigma x^4 \cdot \Sigma x^0 - (\Sigma x^2)^2$	$\Sigma x^6\,\Sigma x^2 - (\Sigma x^4)^2$	$\Sigma x^4 - \Sigma x^2$
1	3	2	2	2	2	0	1 0
2	5	10	34	130	70	144	3 4
3	7	28	196	1,588	588	6,048	6 0
4	9	60	708	9,780	2,772	85,536	11.8
5	11	110	1,958	41,030	9,438	679,536	17 8
6	13	182	4,550	134 342	26,026	3,747,744	25 0
7	15	280	9,352	369,640	61,880	16,039,296	33 4
8	17	408	17 544	893 928	131,784	56,930,688	43 0
9	19	570	30 666	1 956 810	257,754	174,978,144	53 8
10	21	770	50,666	3 956,810	471,086	479,700,144	65 8
11	23	1,012	79,948	7,499,932	814,660	1,198,248,480	79 0
12	25	1,300	121,420	13,471 900	1,345,500	2,770,653,600	93 4
13	27	1,638	178,542	23,125,518	2,137,590	6,002,352,720	109 0
14	29	2,030	255,374	38,184,590	3,284,946	12,298,837 824	125 8
15	31	2,480	356,624	60,965,840	4,904,944	24,014,605,824	143 8
16	33	2,992	487,696	94,520,272	7,141,904	44,957,265,408	163 0
17	35	3,570	654,738	142,795,410	10,170,930	81,097,765,056	183 4
18	37	4,218	864 690	210 819,858	14,202,006	141,549 364,944	205 0
19	39	4,940	1,125,332	304,911,620	19,484,348	239,891,292,576	227 8
20	41	5,740	1,445,332	432,911,620	26,311,012	395,928,108,576	251 8
21	43	6 622	1 834,294	604 443 862	35,023,758	637,992,775,728	277 0
22	45	7,590	2,302,806	831,203,670	46,018,170	1,005,920,381,664	303 4
23	47	8,648	2,862,488	1,127,275,448	59,749 032	1,554,840,524,160	331 0
24	49	9,800	3,526,040	1,509,481,400	76 735,960	2,359,959,638,400	359 8
25	51	11,050	4,307,290	1,997,762,650	97 569 290	3,522,530,138,400	389.8

TABLE TO BE USED WHEN THE NUMBER OF OBSERVATIONS IS EVEN

n	ΣX^0	ΣX^2	ΣX^4	ΣX^6	$\Sigma X^4 \Sigma X^0 - (\Sigma X^2)^2$	$\Sigma X^6 \Sigma X^2 - (\Sigma X^4)^2$	$\Sigma X^4 - \Sigma X^2$
1	2	2	2	2	0	0	1 0
3	4	20	164	1,460	256	2,304	8 2
5	6	70	1,414	32,710	3 584	290,304	20 2
7	8	168	6,216	268 008	21,504	6,386 688	37 0
9	10	330	19,338	1,330 890	84,480	65,235,456	58 6
11	12	572	48,620	4 874 012	256 256	424,030,464	85 0
13	14	910	105,742	14 527 630	652 288	2,038 772,736	116 2
15	16	1,360	206,992	37,308,880	1,462 272	7,894,388,736	152 2
17	18	1,938	374,034	85 584,018	2,976,768	25,960,393 728	193 0
19	20	2,660	634,676	179,675 780	5,617,920	75,123,949 824	238 6
21	22	3,542	1,023,638	351,208,022	9,974,272	196,144,058 880	289 0
23	24	4,600	1,583,320	647,279,800	16,839,680	470,584,857,600	344 2
25	26	5,850	2 364,570	1,135,561,050	27,256,320	1,051,840,857,600	404 2
27	28	7,308	3,427,452	1,910 402,028	42,561,792	2 213,790,808,320	469 0
29	30	8 990	4 842,014	3,100,048,670	64,440 320	4,424,337,967 104	538 6
31	32	10,912	6,689,056	4 875,056,032	94,978,048	8,453,141,250,048	613 0
33	34	13,090	9,060,898	7,457,991,970	136,722,432	15,525,242,320,896	692 2
35	36	15,540	12 062,148	11,134,523,220	192,745,728	27,535,076,464,896	776 2
37	38	18,278	15,810,470	16,265,976,038	266,712,576	47,338,548,401,664	865 0
39	40	21,320	20 437,352	23,303,463,560	362,951,680	79,144,486,327,296	958 6
41	42	24,682	26,088,874	32,803,672,042	486,531,584	129,030,886,752,768	1 057 0
43	44	28,380	32,926,476	45,446,398,140	643,340,544	205,615,957,434,624	1,160.2
45	46	32,430	41,127,726	62,053,929,390	840,170,496	320,919,084,186,624	1 268 2
47	48	36,848	50,887,088	83,612,360,048	1,084,805,120	491,452 517,928,960	1,381 0
49	50	41,650	62,416 690	111,294,934,450	1,386,112,000	739,590,829,286,400	1 498 6

Finally, the sign of G is determined by the direction in which the curve approaches the asymptote $y = 0$, and this may readily be told by inspection But it not infrequently happens that a slight error in one of the observations may be sufficient to give G the wrong sign In this case the limits between which the observations were taken must be changed, or a new value of k must be tried, or the faulty observation must be adjusted by a smoothing formula It is obviously important therefore that some means be provided for determining the sign of G before the values of the coefficients are determined

The condition that G shall be negative is $\dfrac{\Sigma X^3 Y}{\Sigma X Y} > \dfrac{\Sigma X^4}{\Sigma X^2}$ The second term in this inequality may be tabulated in the same way The accompanying tables show the values of the functions

$$\Sigma X^0, \ \Sigma X^2, \ \Sigma X^4, \ \Sigma X^6, \ \Sigma X^4 \Sigma X^0 - (\Sigma X^2)^2,$$
$$\Sigma X^6 \Sigma X^2 - (\Sigma X^4)^2 \text{ and } \Sigma X^4 - \Sigma X^2$$

for all values of n from 0 to 25 when the number of observations is odd and from 0 to 49 when they are even

In the preparation of these tables, my thanks are due to the Zoological Society of San Diego for the use of the facilities afforded by its research department

Joshua L Bailey Jr

THE GENERALIZATION OF STUDENT'S RATIO*

By

HAROLD HOTELLING

The accuracy of an estimate of a normally distributed quantity is judged by reference to its variance, or rather to an estimate of the variance based on the available sample In 1908 "Student" examined the ratio of the mean to the standard deviation of a sample [1] The distribution at which he arrived was obtained in a more rigorous manner in 1925 by R A Fisher,[2] who at the same time showed how to extend the application of the distribution beyond the problem of the significance of means which had been its original object, and applied it to examine regression coefficients and other quantities obtained by least squares, testing not only the deviation of a statistic from a hypothetical value but also the difference between two statistics

Let ξ be any linear function of normally and independently distributed observations of equal variance, and let s be the estimate of the standard error of ξ derived by the method of maximum likelihood If we let t be the ratio to s of the deviation of ξ from its mathematical expectation, Fisher's result is that the probability that t lies between t_1 and t_2 is

*Presented at the meeting of the American Mathematical Society at Berkeley, April 11, 1931

[1] Biometrika, vol 6 (1908), p 1

[2] Applications of Student's Distribution, Metron, vol 5 (1925), p 90

$$(1) \qquad \frac{1}{\sqrt{\pi n}} \qquad \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{t_1}^{t_2} \frac{dt}{(1+t^2/n)^{\frac{n+1}{2}}}$$

where n is the number of degrees of freedom involved in the estimate s

It is easy to see how this result may be extended to cases in which the variances of the observations are not equal but have known ratios and in which, instead of independence among the observations, we have a known system of intercorrelations. Indeed, we have only to replace the observations by a set of linear functions of them which are independently distributed with equal variance By way of further extension beyond the cases discussed by Fisher, it may be remarked that the estimate of variance s^2 may be based on a body of data not involved in the calculation of ξ Thus the accuracy of a physical measurement may be estimated by means of the dispersion among similar measurements on a different quantity.

A generalization of quite a different order is needed to test the simultaneous deviations of several quantities This problem was raised by Karl Pearson in connection with the determination whether two groups of individuals do or do not belong to the same race, measurements of a number of organs or characters having been obtained for all the individuals Several "coefficients of racial likeness" have been suggested by Pearson and by V Romanovsky with a view to such biological uses Romanovsky has made a careful study[1] of the sampling distributions, assuming in each case that the variates are independently and normally

[1]V Romanovsky, On the criteria that two given samples belong to the same normal population (on the different coefficients of racial likeness), Metron, vol. 7 (1928), no 3, pp. 3-46, K Pearson, On the coefficient of racial likeness, Biometrika, vol 18 (1926), pp 105-118

distributed One of Romanovsky's most important results is the exact sampling distribution of L, a constant multiple of the sum of the squares of the values of t for the different variates This distribution function is given by a somewhat complex infinite series For large samples and numerous variates it slowly approximates to the normal form, for 500 individuals, Romanovsky considers that an adequate approach to normality requires that no fewer than 62 characters be measured in each individual When it is remembered that all these characters must be entirely independent, and that it is usually hard to find as many as three independent characters, the difficulties in application will be apparent To avoid these troubles, Romanovsky proposes a new coefficient of racial likeness, H, the average of the ratios of variances in the two samples for the several characters He obtains the exact distribution of H, again as an infinite series, though it approaches normality more rapidly than the distribution of L. But H does not satisfy the need for a comparison between magnitudes of characters, since it concerns only their variabilities

Joint comparisons of correlated variates, and variates of unknown correlations and standard deviations, are required not only for biologic purposes, but in a great variety of subjects The eclipse and comparison star plates used in testing the Einstein deflection of light show deviations in right ascension and in declination, an exact calculation of probability combining the two least-square solutions is desirable. The comparison of the prices of a list of commodities at two times, with a view to discovering whether the changes are more than can reasonably be ascribed to ordinary fluctuation, is a problem dealt with only very crudely by means of index numbers, and is one of many examples of the need for such a coefficient as is now proposed We shall generalize Student's distribution to take account of such cases

We consider p variates $x_1, x_2, \ldots, x_p$, each of which is measured for N individuals, and denote by $X_{i\alpha}$ the value of x_i for the α th individual. Taking first the problem

of the significance of the deviations from a hypothetical set of mean values $m_1, m_2, \ldots, m_p$, we calculate the means $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p$, of the samples, and put

$$\xi_i = (\bar{x}_i - m_i)\sqrt{N}.$$

Then the mean values of the ξ_i will all be zero, and the variances and covariances will be the same as for the corresponding x_i, since the individuals are supposed chosen independently from an infinite population[1] In order to estimate them with the help of the deviations

$$x_i = X_{i\alpha} - \bar{x}_i$$

from the respective means, we call $n = N - 1$ the number of degrees of freedom and take as the estimates of the variances and covariances,

$$a_{ji} = a_{ij} = \frac{1}{n}\sum_{\alpha=1}^{N} x_{i\alpha} x_{j\alpha}$$

We next put

$$a = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix}$$

[1]"Mean Value" is used in the sense of mathematical expectation, the variance of a quantity whose mean value is zero is defined as the expectation of its squares, the covariance of two such quantities is the expectation of their product Thus the correlation of the two in a hypothetical infinite population is the ratio of their covariance to the geometric mean of the variances

$$(3) \qquad A_{ij} = A_{ji} = \frac{\text{cofactor of } a_{ij} \text{ in } a}{a}$$

The measure of simultaneous deviations which we shall employ is

$$(4) \qquad T^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} A_{ij}\,\xi_i\,\xi_j .$$

For a single variate it is natural to take $A_{11} = 1/a_{11}$, then T reduces to t, the ordinary "critical ratio" of a deviation in a mean to its estimated standard error, a ratio which has "Student's distribution," (1) For examining the deviations from zero of two variates x and y,

$$T = \frac{N}{1-r^2}\left\{\frac{\bar{x}^2}{s_1^2} - \frac{2r\bar{x}\bar{y}}{s_1 s_2} + \frac{\bar{y}^2}{s_2^2}\right\},$$

where

$$s_1^2 = \frac{\Sigma(X-\bar{x})^2}{N-1}, \qquad s_2^2 = \frac{\Sigma(Y-\bar{y})^2}{N-1},$$

$$r = \frac{\Sigma(X-\bar{x})(Y-\bar{y})}{\sqrt{\Sigma(X-\bar{x})^2\,\Sigma(Y-\bar{y})^2}}$$

For comparing the means of two samples, one of N_1 and the other of N_2 individuals, we distinguish symbols pertaining to the second sample by primes, and write

$$(5) \qquad \xi_i = \frac{\bar{x}_i - \bar{x}_i'}{\sqrt{1/N_1 + 1/N_2}}$$

$$n = N_1 + N_2 - 2,$$

$$(6) \qquad a = \frac{1}{n}\left[\Sigma(X_{i\alpha} - \bar{x}_i)(X_{j\alpha} - \bar{x}_j) + \Sigma(X_{i\alpha}' - \bar{x}_i')(X_{j\alpha}' - \bar{x}_j')\right]$$

$$= \frac{1}{n}\left[\Sigma X_{i\alpha} X_{j\alpha} - N_1 \bar{x}_i \bar{x}_j + \Sigma X_{i\alpha}' X_{j\alpha}' - N_2 \bar{x}_i' \bar{x}_j'\right]$$

and take as our "coefficients of racial likeness" the value (4) of T^2, in which the ξ_i are calculated from (5) and the A_{ij} from (6) and (3)

Other situations to which the measure T^2 of simultaneous deviations can be applied include comparisons of regression coefficients and slopes of lines of secular trend, comparisons which for single variates have been explained by R A Fisher[1] In each case we deal for each variate with a linear function ξ_i of the observed values, such that the sum of the squares of the coefficients is unity, so that the variance is the same as for a single observation, and such that the expectation of ξ_i is, on the hypothesis to be tested, zero Deviations $x_{i\alpha}$ of the observations from means, or from trend lines or other such estimates, are used to provide the estimated variances and covariances a_{ij} by (2) The number of degrees of freedom n is the difference between the number N of individuals and the number q of independent linear relations which must be satisfied by the quan-

[1]Metron, loc cit, and Statistical Methods for Research Workers, Oliver and Boyd, third edition (1928)

ties x_{i1}, x_{i2}, , x_{iN} on account of their method of derivation For all the variates, these relations and n must be the same.

The general procedure is to set up what may be called normal values $\bar{x}_{i\alpha}$ for the respective $X_{i\alpha}$, putting

$$x_{i\alpha} = X_{i\alpha} - \bar{x}_{i\alpha} \tag{7}$$

The underlying assumption is that $X_{i\alpha}$ is composed of two parts, of which one, $\varepsilon_{i\alpha}$, is normally and independently distributed about zero with variance σ_i^2 which is the same for all the observations on x_i. The other component is determined by the time, place, or other circumstances of the α'th observation in some regular manner, the same for all the variates Denoting this part by $\eta_{i\alpha}$, we have

$$X_{i\alpha} = \eta_{i\alpha}' + \varepsilon_{i\alpha} .$$

Specifically, we take $\eta_{i\alpha}$ to be a linear function, with known coefficients $g_{\alpha s}$, of q unknown parameters ζ_{i1}, , ζ_{iq} where $q < N$

$$\eta_{i\alpha} = \sum_{s=1}^{q} g_{\alpha s} \zeta_{is} \tag{8}$$

Thus in dealing with a secular trend representable by a polynomial in the time, we may take the g's as powers of the time-variable, the ζ's as the coefficients For differences of means, the g's are 0's and 1's, and the ζ's the true means

We estimate the ζ's by minimizing

$$2V_i = \sum_{\alpha=1}^{N} \varepsilon_{i\alpha}^2 = \sum_{\alpha=1}^{N} (X_{i\alpha} - \eta_{i\alpha})^2 \tag{9}$$

Substituting from (8), differentiating with respect to ζ_{is}, and replacing $\eta_{i\alpha}$ by $\bar{x}_{i\alpha}$ for the minimizing value, we obtain

$$(10) \quad \sum_{\alpha=1}^{N} g_{\alpha s}(X_{i\alpha} - \bar{x}_{i\alpha}) = 0, \qquad (s = 1, 2, \cdots, q)$$

or by (7),

$$(11) \quad \sum_{\alpha=1}^{N} g_{\alpha s}\, x_{i\alpha} = 0 \qquad (s = 1, 2, \quad, q)$$

Denoting also the minimizing values of ζ_{is} by z_{is}, we have made from (8),

$$\bar{x}_{i\alpha} = \sum_{s=1}^{q} g_{\alpha s}\, z_{is}$$

Subtracting (8),

$$(12) \qquad \bar{x}_{i\alpha} - \eta_{i\alpha} = \sum_{s=1}^{q} g_{\alpha s}(z_{is} - \zeta_{is})$$

From (9),

$$2V = \sum_{\alpha=1}^{N} \left[(X_{i\alpha} - \bar{x}_{i\alpha}) + (\bar{x}_{i\alpha} - \eta_{i\alpha})\right]^2$$

$$(13) \qquad = \sum_{\alpha=1}^{N} (X_{i\alpha} - \bar{x}_{i\alpha})^2 + 2\sum_{\alpha=1}^{N} (X_{i\alpha} - \bar{x}_{i\alpha})(\bar{x}_{i\alpha} - \eta_{i\alpha})$$

$$+ \sum_{\alpha=1}^{N} (\bar{x}_{i\alpha} - \eta_{i\alpha})^2$$

The middle term, by (12), equals

$$2\sum_{\alpha=1}^{N} \sum_{s=1}^{q} g_{\alpha s}(X_{i\alpha} - \bar{x}_{i\alpha})(z_{is} - \zeta_{is}),$$

this, by (10), is zero. Hence, by (7) and (13),

$$U_i = V_i + W_i\,,$$

where

$$2V_i = \sum_{\alpha=1}^{N} x_{i\alpha}^2$$

$$2W_i = \sum_{\alpha=1}^{N} (\bar{x}_{i\alpha} - \eta_{i\alpha})^2$$

If the q equations (10) be solved for $\bar{x}_{i1}$, $\bar{x}_{i2}$, ..., $\bar{x}_{iN}$, the values of these quantities will be found to be homogeneous linear functions of the observations $X_{i\alpha}$. By (7) therefore, the quantities

$$\bar{x}_{i1},\ \bar{x}_{i2},\ \ldots,\ \bar{x}_{iN}$$

are homogeneous linear functions of the $X_{i\alpha}$. But they are not linearly independent functions, since they are connected by the q relations (11). Hence V is a quadratic form of rank

$$n = N - q\,.$$

Since V_i, by (9), is of rank N, W is of rank q.

This shows that Np new quantities $x'_{i\alpha}$, given by equations of the form

$$x'_{i\alpha} = \sum_{\beta=1}^{N} c_{\alpha\beta} x_{i\beta} = \sum_{\beta=1}^{N} C_{\alpha\beta} X_{i\beta}, \quad (\alpha = 1, 2, \ldots, n)$$

(14)

$$x'_{i\alpha} = \sum_{\beta=1}^{N} c_{\alpha\beta}(\bar{x}_{i\beta} - \eta_{i\beta}) = \sum_{\beta=1}^{N} (C_{\alpha\beta} X_{i\beta} - c_{\alpha\beta}\,\eta_{i\beta}), \quad (\alpha = n+1, \ldots, N)$$

can be found such that

$$2V_i = \sum_{\alpha=1}^{N} x_{i\alpha}^2 = \sum_{\alpha=1}^{N} x'^2_{i\alpha},$$

(15)

$$2W_i = \sum_{\alpha=n+1}^{N} x'^2_{i\alpha},$$

and therefore

(16) $$2U_i = \sum_{\alpha=1}^{n} x'^2_{i\alpha}$$

Substituting (14) in (15) and equating like coefficients,

(17) $$\sum_{\alpha=1}^{n} c_{\alpha\beta}\, c_{\alpha\gamma} = \delta_{\beta\gamma}$$

where $\delta_{\beta\gamma}$ is the Kronecker delta, equal to 1 if $\beta = \gamma$, to 0 if $\beta \neq \gamma$

The coefficients $c_{\alpha\beta}$ depend only on the $g_{\alpha s}$, which have been assumed to be the same for all the p variates Thus (14) may be written

$$x'_{j\alpha} = \sum_{\gamma=1}^{N} c_{\alpha\gamma}\, x_{j\gamma}.$$

Multiplying by (14), summing with respect to α from 1 to n, and using (17),

$$\sum_{\alpha=1}^{n} x'_{i\alpha} x'_{j\alpha} = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N} c_{\alpha\beta}\, c_{\alpha\gamma}\, x_{i\beta}\, x_{j\gamma}$$

(18)

$$= \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N} \delta_{\beta\gamma}\, x_{i\beta}\, x_{j\gamma} = \sum_{\beta=1}^{N} x_{i\beta}\, x_{j\beta}$$

Just as in (2), we define a_{ij} in this generalized case by

$$(19) \qquad a_{ij} = \frac{1}{n}\sum_{\alpha=1}^{N} x_{i\alpha}\, x_{j\alpha}$$

Then by (18),

$$(20) \qquad a_{ij} = \frac{1}{n}\sum_{\alpha=1}^{N} x'_{i\alpha}\, x'_{j\alpha}\,.$$

Of the last equation, (6) is a special case

The random parts $\varepsilon_{i\alpha}$ of the observations on x_i have by hypothesis the distribution

$$\frac{1}{(\sigma_i\sqrt{2\pi})^N}\, e^{-V_i/2\sigma_i^2}\, d\varepsilon_{i1}\, d\varepsilon_{i2} \quad d\varepsilon_{iN}\,,$$

where V_i is given by (9) From what has been shown, it is clear that this may be transformed into

$$\frac{1}{(\sigma_i\sqrt{2\pi})^N}\, e^{-(x_{i1}'^2 + x_{i2}'^2 + \cdots + x_{iN}'^2)/2\sigma_i^2}\, dx'_{i1} \quad dx'_{iN}\,,$$

showing that $x'_{i1}, \ldots, x'_{iN}$ are normally and independently distributed with equal variance σ_i^2

The statistic ξ_i must be independent of the quantities $x'_{i1}, x'_{i2}, \ldots, x'_{in}$ entering into (20), its mean value must be zero, and its variance must be σ_i^2. These conditions are satisfied in the cases which have been mentioned, and are satisfied in general if ξ_i is a linear homogeneous function of $x'_{i,n+1}, \ldots, x'_{iN}$ with the sum of the squares of the coefficients equal to unity

The measure of simultaneous discrepancy is

$$T^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} A_{ij}\,\xi_i\,\xi_j\,,$$

A_{ij} being defined by (3) on the basis of (19) It is evident that

$$(21)\quad T^2 = -\frac{\begin{vmatrix} 0 & \xi_1 & \xi_2 & \cdots & \xi_p \\ \xi_1 & a_{11} & a_{12} & \cdots & a_{1p} \\ \xi_2 & a_{21} & a_{22} & \cdots & a_{2p} \\ \cdot & \cdot & \cdot & & \cdot \\ \xi_p & a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdot & \cdot & & \cdot \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix}}$$

as appears when the numerator is expanded by the first row, and the resulting determinants by their first columns

A most important property of T is that it is an absolute invariant under all homogeneous linear transformations of the variates $x_1, \ldots, x_p$ This may be seen most simply by tensor analysis; for ξ_i is covariant of the first order and A_{ij} is contravariant of the second order

The invariance of T shows that in seeking its sampling distribution we may, without loss of generality, assume that the variates $x_1, \ldots, x_p$ have, in the normal population, zero correlations and equal variances for they may always by a linear transformation be replaced by such variates

Let us now take

$$\xi_i,\ x'_{i1},\ x'_{i2},\ \ldots,\ x'_{in}$$

as rectangular coordinates of a point P_1 in space V_{n+1} of $n+1$ dimensions Since these quantities are normally and independently distributed with equal variance about zero, the probability density for P_i has spherical symmetry about the origin Indefinite repetition of the sampling would result in a globular cluster of representative points for each variate Actually the sample in hand fixes the points P_1, P_2, . , P_p, which may be regarded as taken independently

We shall now show that T is a function of the angle θ between the ξ-axis and the flat space V_p, containing the points P_1, P_2, . , P_p and the origin 0 We shall denote by A the point on the ξ-axis of coordinates 1, 0, 0, , 0, and by V_n the flat space containing the remaining axes Since in V_{n+1} one equation specifies V_n and $n+1-p$ equations V_p, the intersection of V_n and V_p is specified by all these $n+2-p$ equations, and is therefore of $p-1$ dimensions. Call it V_{p-1}

If P_1, P_2, ,, P_p be moved about in V_p, θ will not change, and neither will T, since T is invariant under linear transformations, equivalent to such motions of the P_i Hence T always has the value which it takes if all the lines OP_1, OP_2, , OP_p are perpendicular, with the last $p-1$ of these lines lying in V_{p-1} In this case the angle AOP_1 equals θ Applying to the coordinates of A and of P_1 the formula for the cosine of an angle at the origin of lines to (x_1, x_2,) and (y_1, y_2,), namely,

$$(22) \qquad \cos\theta = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

We obtain

$$\cos\theta = \frac{\xi}{\sqrt{\xi_1^2 + x_{11}'^2 + \quad + x_{1n}'^2}}.$$

Since $x_{11}'^2 + \quad + x_{1n}'^2 = na_{11}$,

it follows that

(23) $\quad n \cot^2 \theta = \xi_1^2 / a_{11}$

The fact that P_2, P_3, $\quad$, P_p lie in V_{p-1}, and therefore in V_n, shows that in this case

$$\xi_2 = \xi_3 = \quad = \xi_p = 0$$

Because OP_1, OP_2, $\quad$, OP_p' are mutually perpendicular, (20) and (22) show that $a_{ij} = 0$ whenever $i \neq j$ Hence, by (21) and (23),

(24) $\quad T = \xi_1 / a_{11} = \sqrt{n} \cot \theta$

By this result the problem of the sampling distribution of T is reduced to that of the angle θ between a line OA in V_{n+1} and the flat space V_p containing p other lines drawn independently through the origin The distribution will be unaffected if we suppose V_p fixed and OA drawn at random, with spherical symmetry for the points A.[1] Let us then, abandoning the coordinates hitherto used, take new axes of rectangular coordinates y_1, y_2, $\quad$, y_{n+1}, of which the first p lie in V_p A unit hypersphere about O is defined in terms of the general-

[1] This geometrical interpretation of T shows its affinity with the multiple correlation coefficient, whose interpretation as the cosine of an angle of a random line with a V_p enabled R A Fisher to obtain its exact distribution (Phil Trans, vol 213B, 1924, p 91, and Proc Roy Soc., vol 121A, 1928, p 654) The omitted steps in Fisher's argument may be supplied with the help of generalized polar coordinates as in the text Other examples of the use of these coordinates in statistics have been given by the author in The Distribution of Correlation Ratios Calculated from Random Data, Proc Nat Acad Sci, vol 11 (1925), p 657, and in The Physical State of Protoplasm, Koninklijke Akademie van Wetenschappen te Amsterdam, verhandlingen, vol 25 (1928), no. 5, pp 28-31

ized latitude-longitude parameters $\phi_1, \ldots, \phi_n$ if we put

$$
\begin{aligned}
y_1 &= \sin\phi_1 \sin\phi_2 \sin\phi_3 \cdots \sin\phi_{p-1} \cos\phi_p \\
y_2 &= \cos\phi_1 \sin\phi_2 \sin\phi_3 \cdots \sin\phi_{p-1} \cos\phi_p \\
y_3 &= \cos\phi_2 \sin\phi_3 \cdots \sin\phi_{p-1} \cos\phi_p \\
y_4 &= \cos\phi_3 \cdots \sin\phi_{p-1} \cos\phi_p \\
&\cdots \\
y_p &= \cos\phi_{p-1} \cos\phi_p \\
y_{p+1} &= \sin\phi_p \cos\phi_{p+1} \\
&\cdots \\
y_n &= \sin\phi_p \sin\phi_{p+1} \cdots \cos\phi_n \\
y_{n+1} &= \sin\phi_p \sin\phi_{p+1} \cdots \sin\phi_n,
\end{aligned}
$$

for the sum of the squares is unity. Since

$$y_{p+1}^2 + \cdots + y_{n+1}^2 = \sin^2\phi_p$$

we have

$$\phi_p = \theta .$$

The element of probability is proportional to the element of generalized area, which is given by

$$\sqrt{D}\, d\phi_1\, d\phi_2 \cdots d\phi_n ,$$

where D is an n-rowed determinant in which the element in the ith row and jth column is

$$\sum_{k=1}^{n+1} \frac{\partial y_k}{\partial \phi_i} \cdot \frac{\partial y_k}{\partial \phi_j}$$

For $i \neq j$, this is zero Of the diagonal elements, the first p-1 contain the factor $\cos^2 \phi_p$, the pth is unity, and the remaining $n-p$ elements contain the factor $\sin^2 \phi_p$ Since ϕ is not otherwise involved, the element of area is the product of

$$\cos^{p-1}\phi_p \sin^{n-p}\phi_p \, d\phi_p$$

by factors independent of ϕ_p The distribution function of θ is obtained by replacing ϕ_p by θ and integrating with respect to the other parameters Since θ lies between 0 and $\pi/2$, we divide by the integral between these limits and obtain for the frequency element,

$$\frac{2\Gamma(\frac{n+1}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{n-p+1}{2})} \cos^{p-1}\theta \sin^{n-p}\theta \, d\theta.$$

Substituting from (24) we have as the distribution of T

$$(25) \quad \frac{2\Gamma(\frac{n+1}{2})}{\Gamma(p/2)\Gamma(\frac{n-p+1}{2})n^{p/2}} \quad \frac{T^{p-1}dT}{(1+\frac{T^2}{n})^{\frac{n+1}{2}}}$$

For $p = 1$ this reduces to the form of Student's distribution given by Fisher and tabulated in the issue of Metron cited, however, as T may be negative as well as positive in this case, Fisher omits the factor 2

For $p = 2$ the distribution becomes

$$\frac{n-1}{n} \quad \frac{T\,dT}{(1+\frac{T^2}{n})^{\frac{n+1}{2}}}.$$

From this it is easy to calculate as the probability that a given value of T will be exceeded by chance,

$$(26) \qquad P = \frac{1}{(1+\frac{T^2}{n})^{\frac{n-1}{2}}}$$

a very convenient expression

The probability integral for higher values of p may be calculated in various ways, the most direct being successive integration by parts, giving a series of terms analogous to (26) to which, if p is odd, is added an integral which may be evaluated with the help of the tables of Student's distribution If p is large, this process is laborious, but other methods are available

The probability integral is reduced to the incomplete beta function if we put

$$x = (1 + T^2/n)^{-1},$$

for then the integral of (25) from T to infinity becomes

$$P = I_x\left(\frac{n-p+1}{2}, \frac{p}{2}\right),$$

the notation being

$$B_x(p,q) = \int_0^x x^{p-1}(1-x)^{q-1}dx,$$

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx,$$

$$I_x(p,q) = \frac{B_x(p,q)}{B(p,q)}$$

Many methods of calculation have been discussed by H E Soper[1] and by V Romanovsky[2] An extensive table of the incomplete beta function being prepared under the supervision of Professor Karl Pearson has not yet been published

Perhaps the most generally useful method now available is

[1]Tracts for Computers, no 7 (1921)

[2]On certain expansions in series of polynomials of incomplete B-functions (in English), Recueil Math de la Soc de Moscou, vol 33 (1926), pp 207-229.

to make the substitution

$$z = \frac{1}{2} \log_e (n-p+1) T^2 - \frac{1}{2} \log_e np,$$

$$n_1 = p$$

$$n_2 = n-p+1,$$

reducing (25) to a form considered by Fisher Table VI in his book, Statistical Methods for Research Workers, gives the values of z which will be exceeded by chance in 5 per cent and in 1 per cent of cases If the value of z obtained from the data is greater than that in Fisher's table, the indication is that the deviations measured are real

If the variances and covariances are known a priori, they are to be used instead of the a_{ij}, the resulting expression T has the well known distribution of χ, with p degrees of freedom For very large samples the estimates of the covariances from the sample are sufficiently accurate to permit the use of the χ distribution for T This is well shown by (25), in which, as n increases, the factor involving T approaches

$$T^{p-1} e^{-T^2/2} dT,$$

which is proportional to the frequency element for χ when χ is put for T

As Pearson pointed out, the labor of calculating χ, which we replace by T, is prohibitive when forty or fifty characters are measured on each individual With two, three, or four characters, however, the labor is very moderate, and the results far more accurate than any attainable with the Pearson coefficient, The great advantage of using T is the simplicity of its distribution, with its complete independence of any correlations among the variates which may exist in the population

To means of a single variate, it is customary to attach a

"probable error," with the assumption that the difference between the true and calculated values is almost certainly less than a certain multiple of the probable error. A more precise way to follow out this assumption would be to adopt some definite level of probability, say $P = .05$, of a greater discrepancy, and to determine from a table of Student's distribution the corresponding value of t, which will depend on n; adding and subtracting the product of this value of t by the estimated standard error would give upper and lower limits between which the true values may with the given degree of confidence be said to lie. With T an exactly analogous procedure may be followed, resulting in the determination of an ellipse or ellipsoid centered at the point ξ_1, ξ_2, . . . , ξ_p. Confidence corresponding to the adopted probability P may then be placed in the proposition that the set of true values is represented by a point within this boundary.

Harold Hotelling

Vol. II No. 4

THE ANNALS

of

MATHEMATICAL STATISTICS

SIX DOLLARS PER ANNUM

NOVEMBER - 1931

PUBLISHED QUARTERLY BY

AMERICAN STATISTICAL ASSOCIATION

Publication Office—Edwards Brothers, Inc, Ann Arbor, Michigan

Business Office—530 Commerce Bldg., New York Univ., New York, N. Y.

Entered as second class matter at the Postoffice at Ann Arbor, Mich., under the Act of March 3rd, 1879.

SYSTEMS OF POLYNOMIALS CONNECTED WITH THE CHARLIER EXPANSIONS AND THE PEARSON DIFFERENTIAL AND DIFFERENCE EQUATIONS*

By

EMANUEL HENRY HILDEBRANDT

INTRODUCTION

The problem of fitting mathematical curves to statistical data has commanded the attention of statisticians and mathematicians for many years The curves referred to the most by English-speaking biometricians and mathematicians are perhaps those developed by Pearson from 1895-1916[1] He showed that a series of curves could be obtained by assigning various values to the parameters in a certain first order differential equation A few years later, Charlier[2], attacking the same question from a differ-

---

*A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the University of Michigan—August, 1931

[1]Karl Pearson, "Mathematical Contributions to the Theory of Evolution," Philosophical Transactions, A, Vol 186 (1895), pp 343-414, also "Supplement to a Memoir on Skew Variation," Phil Trans, Vol 197 (1901), pp 443-456, also "Second Supplement to a Memoir on Skew Variation," Phil Trans, A, Vol 216 (1916), pp 429-457

[2]C V L Charlier, "Ueber das Fehlergesetz," Arkiv for Matematik, Astronomi och Fysik, Vol 2, No 8 (1905), pp 1-9, also "Ueber die Darstellung willkuerlicher Funktionen," Arkiv for Matematik, Astronomi och Fysik, Vol 2, No 20 (1905), pp 1-35

ent angle, showed that any function could probably be approximated by using a certain function and its derivatives in the terms of the series.

$$F(x) = A_0 f(x) + A_1 f'(x) + A_2 f''(x) + \cdots$$

where the A_i are constants

Charlier found that the constants A_n could be formally determined, the n th constant A_n being dependent on the moments of $F(x)$ of order not greater than n He illustrated the method of procedure for the case where $y = f(x)$ was the equation of the normal curve of error, i e one of the Pearson curves In fact, the successive derivatives of this particular function gave rise to a well known system of polynomials, namely the Hermite polynomials, and the coefficients are dependent upon these polynomials also

In recent years, Romanovsky[1] has succeeded in obtaining similar results for the case in which some of the other of the Pearson curves are used as the $f(x)$ in the Gram-Charlier series The successive derivatives of these other special Pearson type curve functions also result in systems of polynomials which bear fundamental relations to each other

It is the object of this investigation to show

(1) That the constants obtained by Charlier for his Type A series can be much more readily obtained by making use of certain existing biorthogonality conditions,

(2) That if the Type A series be generalized to the form

$$F(x) = C_0 Q(x) f_0(x) + C_1 \frac{d}{dx} Q(x) f_1(x) + C_2 \frac{d^2}{dx^2} Q(x) f_2(x) +$$

[1] V Romanovsky, "Generalization of some types of the frequency curves of Professor Pearson," Biometrika, Vol 16 (1924), pp 106-117, also "Sur quelques classes nouvelles de Polynomes orthogonaux," Comptes Rendus de L'Academie des Sciences, Vol 188 (1929), pp. 1023-1025.

where $f_n(x)$ is a polynomial of degree n in x, then the C_n can also be formally determined and depend upon the moments of $F(x)$ of order at most n,

(3) That the form of the polynomials obtained by Charlier and Romanovsky for certain solutions of the Pearson differential equation can be found for any solution of this equation and that the relations existing between polynomials of the same system can also be generalized for the general solution and for the most part obtained without having the explicit form of the solution,

(4) That results analogous to those obtained in (1) and (3) can be derived for the Charlier Type B series and the analogue of Pearson's differential equation, finite differences replacing the derivative

The writer wishes to particularly express his appreciation to Prof H C Carver for the valuable aid he has given both in the stimulating instruction characterized by frankness in indicating unsolved problems in his classes and through direct suggestions in the preparation of this paper

CHAPTER I

POLYNOMIALS CONNECTED WITH THE GRAM-CHARLIER SERIES

1 In the articles entitled "Ueber das Fehlergesetz" and "Ueber die Darstellung willkurlicher Funktionen"[1] Charlier proves the following well known theorem

CHARLIER'S THEOREM FOR SERIES OF TYPE A—*If $F(x)$ is any real valued function of x, which has finite moments of all orders, then $F(x)$ may be formally expressed in terms of another function $f(x)$ and its derivatives as follows*

$$(A) \quad F(x) = A_0 f(x) + A_1 f'(x) + A_2 f''(x) + \quad + A_n f^{(n)}(x) +$$

where $f(x)$ has the following properties

(a) *$f(x)$ and its derivatives are continuous for all real values of x,*

(b) *$f(x)$ and its derivatives vanish for* $x = +\infty$ *and* $-\infty$

(c) $\lim_{x \to \pm\infty} x^m f^{(n)}(x) \quad 0$ *for all m and n,*

(d) $\int_{-\infty}^{+\infty} f(x)\, dx \neq 0$

The conditions (c) and (d) are not given in Charlier's articles, but an eaxmination of the proof shows that he assumes implicitly that they are satisfied $f(x) = \frac{x}{1+x^2}$ satisfies (a) and (b) without satisfying (c) and (d)

In the first section of the latter paper, Charlier determines the constants $A_0, A_1, A_2, \quad , A_n,$ He takes the series (A), multiplies it successively by 1, x, x^2, , and integrates each result between the limits $-\infty$ to $+\infty$ The fol-

[1] C V L Charlier, loc cit

lowing equations result

$$\int_{-\infty}^{+\infty} F(x)dx = A_o \int_{-\infty}^{+\infty} f(x)dx$$

$$\int_{-\infty}^{+\infty} x F(x)dx = A_o \int_{-\infty}^{+\infty} x f(x)dx + A_1 \int_{\infty}^{+\infty} x f'(x)\,dx$$

$$\int_{-\infty}^{+\infty} x^2 F(x)dx = A_o \int_{-\infty}^{+\infty} x^2 f(x)dx + A_1 \int_{\infty}^{+\infty} x^2 f'(x)dx + A_2 \int_{\infty}^{+\infty} x^2 f''(x)dx$$

Each of these equations contain a finite number of terms and the constants $A_o, A_1, A_2,$ may readily be determined by solving them In fact we find that any constant A_n may be expressed as

$$A_n = \int_{-\infty}^{+\infty} P_n(x) F(x) dx$$

where $P_n(x)$ is a polynomial in x of degree not greater than n An analysis of the underlying facts reveals that what Charlier has actually done is to show that under the conditions listed in the theorem there exists a uniquely determined set of polynomials $P_o(x)$, $P_1(x)$, , $P_n(x)$, , $P_n(x)$ at most of degree n, biorthogonal to the set of derivatives or functions of $f(x)$, i e satisfy the biorthogonality conditions

$$\int_{\infty}^{+\infty} P_n(x) f^{(m)}(x) dx = 0 \text{ for } m \neq n$$

$$= 1 \text{ for } m = n$$

Further a study of the coefficients of these polynomials shows that

$$\frac{d\,P_n(x)}{dx} = - P_{n-1}(x),$$

i e we have the following theorem

THEOREM: *If $f(x)$ satisfy the conditions* (a), (b), (c), *and* (d) *of Charlier's theorem for series* (A) *and if $P_0(x)$, $P_1(x)$, . $P_n(x)$. is the system of polynomials in x, $P_n(x)$ of degree at most n, which is biorthogonal to $f(x)$ and its derivatives, i. e. satisfies the conditions*

$$\int_{-\infty}^{+\infty} P_n(x)\, f^{(m)}(x)\,dx = 0 \text{ for } m \neq n$$
$$= 1 \text{ for } m = n$$

then

$$\frac{dP_n(x)}{dx} = -P_{n-1}(x)$$

This can readily be shown to be true directly from a use of the biorthogonal property. For integrating by parts we obtain:

$$\int_{-\infty}^{+\infty} P_n(x) f^{(m)}(x)\,dx = P_n(x) f^{(m-1)}(x) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} P_n'(x)\, f^{(m-1)}(x)\,dx.$$

The first half of the right hand side of this equation vanishes due to condition (c) of Charlier's theorem for series (A). For the second half we have

$$-\int_{-\infty}^{+\infty} P_n'(x) f^{(m-1)}(x)\,dx = 0 \text{ for } m \neq n$$
$$= 1 \text{ for } m = n$$

But we know that

$$+\int_{-\infty}^{+\infty} P_{n-1}(x) f^{(m-1)}(x)\,dx = 0 \text{ for } m \neq n$$
$$= 1 \text{ for } m = n$$

determines uniquely the polynomials $P_{n-1}(x)$ It follows that

$$dP_n(x)/dx = -P_{n-1}(x)$$

A corollary to this last theorem may be stated as follows

COROLLARY

$$If \int_{\infty}^{+\infty} P_n(x)\, f^{(m)}(x)\, dx = 0 \qquad for\ m \neq n$$

$$= a_n \qquad for\ m = n$$

$a_i \neq 0$ $(i = 0, 1, 2, \quad)$, *then*

$$d P_n(x)/dx = -\frac{a_n}{a_{n-1}} P_{n-1}(x)$$

The proof is similar to the one just given Integration by parts gives the following result

$$-\int_{-\infty}^{+\infty} f^{(m-1)}(x)\, P_n'(x)\, dx = 0 \qquad \text{for } m \neq n$$

$$= a_n \qquad \text{for } m = n$$

But we know that

$$\int_{-\infty}^{+\infty} f^{(m-1)}(x)\, P_{n-1}(x)\, dx = 0 \qquad \text{for } m \neq n$$

$$= a_{n-1} \qquad \text{for } m = n$$

Therefore we may conclude that

$$-\frac{1}{a_n} \frac{d P_n(x)}{dx} = \frac{1}{a_{n-1}} P_{n-1}(x)$$

or

$$\frac{d P_n(x)}{dx} = -\frac{a_n}{a_{n-1}} P_{n-1}(x)$$

An illustration of this corollary is the case of the well known Hermite polynomials which are involved in Charlier's first paper [1] These satisfy the conditions

[1] C V L Charlier, loc cit Charlier uses as $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-b)^2}{2\sigma^2}}$ In this paper we shall use the simpler basic function e^{-x^2}

$$\int_{-\infty}^{+\infty} H_m(x) H_n(x) e^{-x^2} dx = 0 \qquad \text{for } m \neq n$$

$$= 2^n n! \sqrt{\pi} \qquad \text{for } m = n$$

and

$$H_n(x) e^{-x^2} = (-1)^n d^n (e^{-x^2})/dx^n$$

Hence

$$\int_{-\infty}^{+\infty} H_m(x) d^n (e^{-x^2})/dx^n \quad dx = 0 \qquad \text{for } m \neq n$$

$$= (-2)^n n! \sqrt{\pi} \quad \text{for } m = n$$

If then $f(x) = e^{-x^2}$ and $a_n = (-2)^n n! \sqrt{\pi}$ our corollary applies, i e we have

$$dH_n(x)/dx = 2nH_{n-1}(x)$$

We might further observe that if $a_n = (-1)^n n!$ then the polynomials $P_n(x)$ form a system of Appell polynomials[1] satisfying the relation

$$dP_n(x)/dx = nP_{n-1}(x)$$

the n th polynomial being the coefficient of $h^n/n!$ in the expansion of $a(h) e^{hx}$ where

$$a(h) = \alpha_0 + \frac{h}{1!} \alpha_1 + \frac{h^2}{2!} \alpha_2 + \qquad + \frac{h^n}{n!} \alpha_n +$$

The fact that differentiation of the n th pol nomial results in the negative of the $(n-1)$th polynomial, shows that the n th polynomial may be obtained by integrating the $(n-1)$ th one,

[1] M P Appell, "Sur une classe de Polynomes," Annales Scientifiques de L'Ecole Normale Superioure, Vol IX, series 2 (1880), pp 119-120

which will consequently determine all of the terms of the n th polynomial except the constant This constant may be found from any of the conditions of biorthogonality The simplest of these conditions is

$$\int_{-\infty}^{+\infty} P_n(x)\, f(x)\, dx = 0$$

Setting

$$P_n(x) = -\int_0^x P_{n-1}(x)\, dx + c$$

gives
$$\int_{-\infty}^{+\infty} \left[-\int_0^x P_{n-1}(x)\, dx + c \right] f(x)\, dx = 0$$

and so
$$c = \frac{\int_{-\infty}^{+\infty} \left[\int_0^x P_{n-1}(x)\, dx \right] f(x)\, dx}{\int_{-\infty}^{+\infty} f(x) dx}$$

so that
$$P_n(x) = -\int_0^x P_{n-1}(x) dx + \frac{\int_{-\infty}^{+\infty} \left[\int_0^x P_{n-1}(x)\, dx \right] f(x) dx}{\int_{-\infty}^{+\infty} f(x) dx}$$

This gives a very simple and elegant method of writing down successively the polynomials associated with any function $f(x)$ satisfying the conditions of the theorem

Using the Charlier notation

$$\lambda_n = \frac{\int_{-\infty}^{\infty} x^n f(x) dx}{n!}$$

and observing that $P_0(x) = 1/\lambda_0$, we obtain the following

polynomials

$$P_1(x) = -\int_0^x P_0(x)dx + \frac{\int_{-\infty}^{+\infty}\left[\int_0^x P_0(x)dx\right] f(x)dx}{\int_{-\infty}^{+\infty} f(x)dx}$$

$$= -\frac{x}{\lambda_0} + \frac{\lambda_1}{\lambda_0^2},$$

$$P_2(x) = -\int_0^x P_1(x)dx + \frac{\int_{-\infty}^{+\infty}\left[\int_0^x P_1(x)dx\right] f(x)dx}{\int_{-\infty}^{+\infty} f(x)dx}$$

$$= \frac{x^2}{\underline{|2}\,\lambda_0} - \frac{\lambda_1 x}{\lambda_0^2} + \frac{\lambda_1^2}{\lambda_0^3} - \frac{\lambda_2}{\lambda_0^2},$$

$$P_3(x) = -\int_0^x P_2(x)dx + \frac{\int_{-\infty}^{\infty}\left[\int_0^x P_2(x)dx\right] f(x)dx}{\int_{-\infty}^{+\infty} f(x)dx}$$

$$= -\frac{x^3}{\underline{|3}\,\lambda_0} + \frac{\lambda_1 x^2}{\underline{|2}\,\lambda_0^2} - \frac{\lambda_1^2 x}{\lambda_0^3} + \frac{\lambda_2 x}{\lambda_0^2} + \frac{\lambda_1^3}{\lambda_0^4} - \frac{2\lambda_2\lambda_1}{\lambda_0^3} + \frac{\lambda_3}{\lambda_0^2},$$

. . . .

2 Just as the Hermite polynomials, based as they are on the derivatives of e^{-x^2}, are the starting point for expansions of the Gram-Charlier type and for the theorem just considered, so the Laguerre polynomials defined by $d^n(a+bx)^n e^{-x}/dx^n$ suggest an expansion of the type

$$F(x) = C_0 f_0(x)\varphi(x) + C_1 \frac{d}{dx} f_1(x)\varphi(x) - C_2 \frac{d^2}{dx^2} f_2(x)\varphi(x) + \cdot$$

where $f_n(x)$ is a polynomial in x As a matter of fact we can state the following theorem

THEOREM If $\phi(x)$ *is a function such that*

(1) $\phi(x)$ *and all its derivatives are continuous for all real values of* x,

(2) $\phi(x)$ *and its derivatives are zero at* $x = +\infty$ *and* $-\infty$,

(3) $\lim_{x \to \pm\infty} x^n \phi^m(x) = 0$,

(4) $\{f_n(x)\}$ *is a sequence of polynomials in* x *such that* $\int_{-\infty}^{+\infty} f_n(x)\phi(x) \neq 0$,

then there exists a unique sequence of polynomials $P_m(x)$, $P_m(x)$ *at most of degree* m, *such that*

$$\int_{-\infty}^{+\infty} P_m(x) \frac{d^n}{dx^n} f_n(x) \phi(x)\, dx = 0 \qquad \text{for } m \neq n$$

$$= 1 \qquad \text{for } m = n$$

If $f_n(x)$ *is at most of degree* n, *then the determination of* $P_m(x)$ *depends at most upon the moments of* ϕ *of order* n [1]

The method of proof is modelled on Charlier's proof for the preceding case By substituting in the n th integration by parts formula

$$\int u(x) v^{n+1}(x) dx = u v^{(n)} - u' v^{(n-1)}$$
$$+ u'' v^{(n-2)} - \quad + (-1)^n u^{(n)} v$$
$$+ (-1)^{n+1} \int u^{(n+1)}(x) v(x)\, dx ,$$

we have

[1] The Laguerre polynomials are not a special case of this because there the interval of integration is $-a/b$ to $+\infty$

$$\int_{-\infty}^{+\infty} P_m(x)\frac{d^n}{dx^n} f_n(x)\phi(x)dx = \left\{P_m(x)\frac{d^{n-1}}{dx^{n-1}} f_n(x)\phi(x) - \frac{dP_m(x)}{dx}\frac{d^{n-2}}{dx^{n-2}} f_n(x)\phi(x)\right.$$

$$+\frac{d^2}{dx^2}P_m(x)\left[\frac{d^{n-3}}{dx^{n-3}}\ f_n\ (x)\phi(x)\right]$$

$$+\quad +(-1)^{n-1}\left.\left[\frac{d^{n-1}}{dx^{n-1}} P_m(x)\right] f_n\phi(x)\right\}_{-\infty}^{+\infty}$$

$$+(-1)^n\int_{-\infty}^{+\infty}\left[\frac{d^n}{dx^n} P_m(x)\right] f_n(x)\phi(x)dx$$

$$=(-1)^n\int_{-\infty}^{+\infty}\left[\frac{d^n}{dx^n} P_m(x)\right] f_n(x)\ \phi(x)\ dx$$

because of conditions (2) and (3) on $\phi(x)$ As a consequence, if $n > m$ then $\frac{d^n}{dx^n} P_m(x) = 0$, so that for $n > m$

$$\int_{-\infty}^{+\infty} P_m(x)\ \frac{d^n}{dx^n}\ f_n(x)\ \phi(x)dx = 0$$

that is to say $P_m(x)$ is orthogonal to $\frac{d^n}{dx^n}\ f_n\ (x)\phi(x)$ provided $n > m$ Hence $P_n(x)$ must satisfy only the following $n+1$ equations

$$\int_{-\infty}^{+\infty} P_m(x)\frac{d^n}{dx^n}\ f_n(x)\phi(x)dx = 0$$

$$\int_{-\infty}^{+\infty} P_n(x)\frac{d}{dx}\ f_1(x)\phi(x)dx = (-1)\int_{-\infty}^{+\infty}\frac{d\,P_n(x)}{dx} f_1\ (x)\ \phi\ (x)dx = 0$$

$$\int_{-\infty}^{+\infty} P_n(x)\frac{d^2}{dx^2}\ f_2(x)\phi(x)dx = (-1)^2\int_{-\infty}^{+\infty}\frac{d^2P_n(x)}{dx^2}\ f_2(x)\ \phi\ (x)dx = 0$$

$$\cdot \qquad\qquad\qquad\qquad \cdot \qquad \cdot \qquad \cdot$$

$$\int_{-\infty}^{+\infty} P_n(x)\frac{d^n}{dx^n}\ f_n(x)\phi(x)dx = (-1)^n\int_{-\infty}^{+\infty}\frac{d^n}{dx^n}\ P_n(x) f_n(x)\phi(x)dx - 1$$

Replacing now $P_n(x)$ by $a_0 + a_1x + a_2x^2 + \quad + a_nx^n$ gives us the system of algebraic equations to be satisfied by $a_0, a_1, \quad\cdot\quad, a_n$, viz :

$$a_o\int_{-\infty}^{+\infty} f_o(x)\phi(x)dx + a_1\int_{-\infty}^{+\infty} x f_o(x)\phi(x)dx$$

$$+ a_2\int_{-\infty}^{+\infty} x^2 f_o(x)\phi(x)dx + \qquad + a_n\int_{-\infty}^{+\infty} x^n f_o(x)\phi(x)dx = 0$$

$$a_1\int_{-\infty}^{+\infty} f_1(x)\phi(x)dx + 2a_2\int_{-\infty}^{+\infty} x f_1(x)\phi(x)dx + \quad + na_n\int_{-\infty}^{+\infty} x^{n-1} f_1(x)\phi(x)dx = 0$$

$$2a_2\int_{-\infty}^{+\infty} f_2(x)\phi(x)dx + \quad + n(n-1)a_n\int_{-\infty}^{+\infty} x^{n-2} f_2(x)\phi(x)dx = 0$$

$$(n-2)!a_{n-2}\int_{-\infty}^{+\infty} f_{n-2}(x)\phi(x)dx + \frac{(n-1)!}{1!}a_{n-1}\int_{-\infty}^{+\infty} x f_{n-2}(x)\phi(x)dx$$

$$+\frac{n!}{2!}a_n\int_{-\infty}^{+\infty} x^2 f_{n-2}(x)\phi(x)dx = 0$$

$$(n-1)!a_{n-1}\int_{-\infty}^{+\infty} f_{n-1}(x)\phi(x)dx + \frac{n!}{1!}a_n\int_{-\infty}^{+\infty} x f_{n-1}(x)\phi(x)dx = 0$$

$$(-1)^n n! a_n\int_{-\infty}^{+\infty} f_n(x)\phi(x)dx = 1$$

We have here a unique determination of a_n if the determinant of the coefficients is $\neq 0$ This is true since the determinant $\Delta = (-1)^n (\int f_o\phi)(\int f_1\phi) \quad (\int f_n\phi)$ is $\neq 0$ because of the condition (4) on ϕ If $f_n(x)$ is at most of degree n, it is obvious that the determination of the $P_n(x)$ resulting from the coefficients a_n depends at most upon the moments of ϕ of order n

The first three polynomials of the type considered in the last theorem have the following form, the limits of integration being $-\infty$ and $+\infty$ in each case

$$P_1(x) = \frac{\int x\phi(x)\,dx}{\int f_1(x)\phi(x)dx\int\phi(x)dx} - \frac{x}{\int f_1(x)\phi(x)dx}$$

$$= \frac{1}{\int f_1(x)\phi(x)dx}\left[\frac{\int x\phi(x)dx}{\int\phi(x)dx} - x\right],$$

$$P_2(x) = \frac{\int x f_1(x)\phi(x)dx\int x\phi(x)dx}{\int f_2(x)\phi(x)dx\int f_1(x)\phi(x)dx\int\phi(x)dx} - \frac{\int x^2\phi(x)dx}{2!\int f_2(x)\phi(x)dx\int\phi(x)dx}$$

$$-\frac{x\int x f_1(x)\phi(x)dx}{\int f_2(x)\phi(x)dx\int f_1(x)\phi(x)dx}+\frac{x^2}{2!\int f_2(x)\phi(x)dx},$$

$$P_3(x)=\frac{\int x f_2(x)\phi(x)dx\int x f_1(x)\phi(x)dx\int x\phi(x)dx}{\int f_3(x)\phi(x)dx\int f_2(x)\phi(x)dx\int f_1(x)\phi(x)dx\int\phi(x)dx}$$

$$-\frac{\int x^2 f_1(x)\phi(x)dx\int x\phi(x)dx}{2!\int f_3(x)\phi(x)dx\int f_1(x)\phi(x)dx\int\phi(x)dx}$$

$$-\frac{\int x f_2(x)\phi(x)dx\int x^2\phi(x)dx}{2!\int f_3(x)\phi(x)dx\int f_2(x)\phi(x)dx\int\phi(x)dx}$$

$$+\frac{\int x^3\phi(x)dx}{3!\int f_3(x)\phi(x)dx\int\phi(x)dx}-\frac{x\int x f_2(x)\phi(x)dx\int x f_1(x)\phi(x)dx}{\int f_3(x)\phi(x)dx\int f_2(x)\phi(x)dx\int f_1(x)\phi(x)dx}$$

$$+\frac{x\int x^2 f_1(x)\phi(x)dx}{2!\int f_3(x)\phi(x)dx\int f_1(x)\phi(x)dx}+\frac{x^2\int x f_2(x)\phi(x)dx}{2!\int f_3(x)\phi(x)dx\int f_2(x)\phi(x)dx}$$

$$-\frac{x^3}{3!\int f_3(x)\phi(x)dx}$$

CHAPTER II

POLYNOMIALS CONNECTED WITH PEARSON'S DIFFERENTIAL EQUATION

1 In the work in mathematical statistics a large number of the problems that require study involve data properly classified into groups and about which further information is sought This data is often classified to form a frequency distribution The frequency distribution when grouped may appear to lie on a certain curve If it can be shown that this curve is a mathematical curve, i e one for which we are able to set up an equation, then this frequency distribution can be readily examined and studied

There are very few frequency distributions which actually conform to known mathematical equations However, there are certain curves which seem to lend themselves much better to statistical manipulations than others Among the most commonly used of these are the so called Pearson type curves Pearson[1] showed in a series of three articles how he obtained the equations of twelve distinct curves and this was done by considering the differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2}$$

and solving it, after assigning particular values to the parameters a_0 , a_1 , b_0 , b_1 , and b_2 The equations of these curves and the differential equations from which they were derived are as follows

[1]Karl Pearson, loc cit

Type	Equation	Differential Equation
I	$y = y_0 \left(1 + \frac{x}{a}\right)^{\nu a} \left(1 - \frac{x}{b}\right)^{\nu b}$	$\frac{dy}{dx} = \frac{\nu(a+b)x}{(a+x)(b-x)} y$
II	$y = y_0 \left(1 - \frac{x^2}{a^2}\right)^{m}$	$\frac{dy}{dx} = \frac{-2mx}{a^2 - x^2} y$
III	$y = y_0 \left(1 + \frac{x}{a}\right)^{\nu a} e^{-\nu x}$	$\frac{dy}{dx} = \frac{-\nu x}{a+x} y$
IV	$y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m} e^{-\nu \arctan \frac{x}{a}}$	$\frac{dy}{dx} = \frac{-2mx - \nu a}{a^2 + x^2} y$
V	$y = y_0 \, x^{-h} e^{-\frac{\nu}{x}}$	$\frac{dy}{dx} = \frac{\nu - hx}{x^2} y$
VI	$y = y_0 \, (x-a)^{q} x^{-h}$	$\frac{dy}{dx} = \frac{ha + (q-h)x}{x^2 - ax} y$
VII	$y = y_0 \, e^{-\frac{x^2}{2\sigma^2}}$	$\frac{dy}{dx} = \frac{-x}{\sigma^2} y$
VIII	$y = y_0 \left(1 + \frac{x}{a}\right)^{-m}$	$\frac{dy}{dx} = \frac{-m}{a+x} y$
IX	$y = y_0 \left(1 + \frac{x}{a}\right)^{m}$	$\frac{dy}{dx} = \frac{m}{a+x} y$
X	$y = \frac{n}{\sigma} e^{\pm \frac{x}{\sigma}}$	$\frac{dy}{dx} = \pm \frac{1}{\sigma} y$
XI	$y = y_0 \, x^{-m}$	$\frac{dy}{dx} = \frac{-m}{x} y$
XII	$y = y_0 \left(\frac{a_1 + x}{a_2 - x}\right)^{p}$	$\frac{dy}{dx} = \frac{h(a_1 + a_2 + 2x)}{(a_1 + x)(a_2 + x)} y$

The curves most widely used are the normal curve of error, which Pearson calls Type VII, and the Type III curve

Suppose a Pearson curve $f(x)$ has been found which seems to fit a given distribution fairly well The question may well be asked Is it possible by means of analytic methods to approach even nearer to the given distribution? For example, would it be possible to use this approximate function as the $f(x)$ in the Charlier series (A) and thus obtain a closer approximation to the observed frequency function

Charlier in his paper "Ueber die Darstellung willkurlicher Functionen"[1] considered this question for $\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-b)^2}{2\sigma^2}}$, i e the normal curve of error He showed that using this $\varphi(x)$ reduced the series (A) to the form

$$(A') \quad F(x) = a_0\varphi(x) + a_3\varphi^{(3)}(x) + a_4\varphi^{(4)}(x) + \quad + a_n\varphi^{(n)}(x) +$$

the first and second derivative terms vanishing due to the proper choice of constants This series (A′) is frequently referred to as the Gram-Charlier Type A series It is worthwhile to note that this $\varphi(x)$ is the same one whose derivatives we found in the first chapter resulted in the Hermite polynomials These polynomials have the following interesting properties[2]

$$(1) \quad dH_n(x)/dx = 2nH_{n-1}(x)$$

$$(2) \quad H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0$$

$$(3) \quad H_n''(x) - 2xH_n'(x) + 2nH_n(x) = 0$$

The first of these relations shows that the derivative of any Hermite polynomial corresponds to the preceding polynomial multi-

---

[1] C V L Charlier, loc cit

[2] R Courant and D Hilbert, Methoden der Mathematischen Physik, 1, pp 76

plied by $2n$ The second equation is a recurrence relation between the $(n+1)$ th, n th and $(n-1)$ th polynomials, while the third relation is a differential equation of the second order involving only the n th polynomial

The use of the equations of the other Pearson type curves as the $f(x)$ in the original Charlier series has in recent years been studied by Romanovsky In the first[1] of two articles, he discusses the Pearson Type I, II and III curves as well as the Type VII—the normal curve referred to in the last paragraph Just as the normal curve of error requires the use of the Hermite polynomials, he found that the Type I curve and Type II, which is a special case of Type I, involved the Jacobi polynomials

$$G_n(h,q,x)=\frac{x^{1-q}(1-x)^{q-h}}{q(q+1)\cdots(q+n-1)}\frac{d^n}{dx^n}\left[x^{q+n-1}(1-x)^{h+n-q}\right]$$

The n 'th Jacobi polynomial satisfies the second order differential equation[2]

$$x(1-x)G_n''(x)+[q-(p+1)x]G_n'(x)+(p+n)nG_n(x)=0$$

which corresponds to property (3) mentioned for the Hermite polynomials above The Type III curve involves the Laguerre polynomials[3] defined by

$$L_n(x)=e^x\frac{d^n}{dx^n}(x^ne^{-x})$$

and these in turn satisfy the recurrence relation

[1] V Romanovsky "Generalization of some types of the frequency curves of Professor Pearson" op cit pp 106 117

[2] R Courant and D Hilbert, op cit, Vol I, p 75

[3] R Courant and D. Hilbert, op cit, pp 77-78

$$L_{n+1}(x) - (2n+1-x)L_n(x) + n^2 L_{n-1}(x) = 0$$

and the differential equation

$$L_n'(x) - n L_{n-1}'(x) = -n L_{n-1}(x)$$

In the second article[2], Romanovsky reviews the cases of the Type IV, V and VI curves The generalization of the Type IV curve gives the polynomial

$$P_n(m,x) = (a^2+x^2)^m e^{\nu\theta} \frac{d^n}{dx^n}\left[(a^2+x^2)^{-m+n} e^{-\nu\theta}\right]$$

where $\theta = \text{arc tan } x/a$ These polynomials possess properties similar to the other polynomials mentioned, viz

$$P_{n+1}(n+1,x) = \left[2(n+1-m)x - \nu a\right] P_n(n,x)$$

$$+ 2n\left[n+1-m\right](a^2+x^2)P_{n-1}(n\ x)$$

and

$$(a^2+x^2)P_n''(n\ x) + \left[2(1-m)x - \nu a\right]$$

$$P_n'(n,x) - n(n+1-2m)P_n(n,x) = 0$$

Similarly for the Type V curve he finds the polynomials

$$P_n(h,x) = x^h e^{\frac{\nu}{x}} \frac{d^n}{dx^n}(x^{-h+2n} e^{-\frac{\nu}{x}})$$

Also the relations

---

[2] V Romanovsky, "Sur quelques Classes nouvells de Polynomes othogonaux,' loc cit

$$P_{n+1}(n+1, x) = \left[(2n+2-p)x+\gamma\right]P_n(n,x) + n(2n+2-p)x^2 P_{n-1}(n,x)$$

and

$$x^2 P_n''(n,x) + \left[x(2-p)+\nu\right]P_n'(n,x) - n(n+1-p)\, P_n(n,x) = 0$$

hold.

Finally for the Type VI curve Romanovsky gets the polynomials

$$P_n(-h\, q, x) = (x-a)^{-q} x^{h} \frac{d^n}{dx^n}\left[(x-a)^{q+n} x^{-h+n}\right]$$

and the relations:

$$P_{n+1}(n+1, x) = \left[(-p+1)(x-a)+(q+1)x\right] P_n(n,x) + x(x-a)P_n'(n,x),$$

$$x(x-a)P_n''(n,x) + \left[(-p+1)(x-a)+(q+1)x\right] P_n'(n,x) - n(n+1+q\; p)P_n(n,x) = 0$$

We note, therefore, that if a solution of the Pearson differential equation is used as the generating function $f(x)$ in the Gram-Charlier series, that a distinct set of polynomials results in each case and that these polynomials satisfy certain recurrence relations and differential equations. These properties are not found in the case of functions such as sech x and sech$^m x$, which were discussed as generating functions by Charlier[1] and by Roa[2] respectively The successive derivatives of the

[1] C V L Charlier, "Ueber die Darstellung willkurlicher Funktionen," loc cit, pp 18 22

[2] Emeterio Roa, "A Number of new generating Functions with Applications to Statistics," Doctor's Thesis, University of Michigan, 1923

sech x do not result in polynomials such as the Hermite or Jacobi ones

Since the generalization of the solutions of the Pearson curves leads to distinct sets of polynomials and since these polynomials satisfy certain fundamental relations, we are led to inquire whether these polynomials are not special cases of a general polynomial and may be obtained from it by specializing the coefficients and further whether such general polynomials, if they do exist, will satisfy certain recurrence relations and differential equations These problems are among those which we shall consider in this chapter

2 In order that we may develop the generalized polynomials let us consider the Pearson differential equation where the numerator is of the first and the denominator of the second degree, i e

$$\frac{1}{y}\frac{dy}{dx} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2}$$

For convenience we shall denote the numerator by N and the denominator by D We then have the following theorem

THEOREM *If y is a non-identically zero solution of*

$$(1) \qquad \frac{dy}{dx} = \frac{N}{D} y$$

then $\frac{D^n}{y}\frac{d^n y}{dx^n}$ *is a polynomial of degree at most n*

The proof will proceed by mathematical induction It is obvious that the theorem holds for $n = 1$, $P_1(x)$ being N Since it is true that

$$D\frac{dy}{dx} = Ny$$

we obtain by differentiation

$$D\frac{d^2y}{dx^2} + D'\frac{dy}{dx} = N\frac{dy}{dx} + N'y ,$$

or using (1) and multiplying the equation through by D we get

$$D^2 \frac{d^2y}{dx^2} = (N^2 - ND' + N'D)y$$

Since D' is linear and N' is a constant, it is obvious that $(N^2 - ND + N'D)$ is at most of degree 2

Assume then that the statement holds for $m \leq n$ and we have

$$D^n \frac{d^n y}{dx^n} = P_n(x)y \tag{2}$$

Differentiation gives

$$nD^{n-1}D'\frac{d^n y}{dx^n} + D^n \frac{d^{n+1}}{dx^{n+1}} = P_n(x)\frac{dy}{dx} + \frac{dP_n(x)}{dx}y$$

Multiplying through by D we get

$$D^{n+1}\frac{d^{n+1}}{dx^{n+1}} = DP_n(x)\frac{dy}{dx} - nD^n D'\frac{d^n y}{dx^n} + D\frac{dP_n(x)}{dx}y,$$

and using (1) and (2), we have

$$\begin{aligned} P_{n+1}(x)\,y &= NP_n(x)y - nD'P_n(x)y + D\frac{dP_n(x)}{dx}y \\ &= \left[NP_n(x) - nD'P_n(x) + D\frac{dP_n(x)}{dx}\right]y \end{aligned}$$

The coefficient of y is obviously a polynomial of degree at most $n+1$. Incidentally we have derived the relation:

$$P_{n+1}(x) = P_n(x)(N - nD') + D\frac{dP_n(x)}{dx} \tag{I}$$

an equation which gives the $(n+1)$ th polynomial in terms of the n th polynomial and its first derivative $P_n'(x)$

3 More generally we have

THEOREM *If y is a non-identically zero solution of* (1), *then*

$$\frac{1}{y} D^{n-k} \frac{d^n}{dx^n} D^k y$$

is a polynomial $P_n(k,x)$, $P_n(k,x)$ *is at most of degree* n *in* x *In particular if* $k=n$, *we have that*

$$\frac{1}{y}\frac{d^n}{dx^n}D^n y$$

is a polynomial in x *of degree at most* n

This theorem can be proved directly following the lines of the preceding theorem, but it is simpler to obtain it as an immediate consequence of this theorem and the following lemma

LEMMA *If* y *satisfy the differential equation* (1) *then* $D^k y$, *where* k *is any real number, satisfies a differential equation of the same type, viz*

$$\frac{d}{dx}(D^k y) = \frac{N+kD'}{D}D^k y$$

Let $u = D^k y$

Then logarithmic differentiation gives at once

$$\frac{1}{u}\frac{du}{dx} = k\frac{D'}{D}+\frac{1}{y}\frac{dy}{dx} = \frac{N+kD'}{D}$$

It follows from this lemma that any result which we derive concerning the polynomials $P_n(x)=\frac{1}{y}D^n\frac{d^n y}{dx^n}$ where y satisfies $D\,dy/dx = Ny$, is immediately extensible to the polynomials $P_n(k,x)=\frac{1}{y}D^{n-k}\frac{d^n}{dx^n}D^k y$ by replacing N by $N+kD'$ In particular relation (I) becomes

$$(\mathrm{I}_k)\quad P_{n+1}(k+1,x)=\left[N+(k-n+1)D'\right]P_n(k+1,x)+D\frac{dP_n(k+1,x)}{dx}$$

which for $k=n$ reduces to

$$(\mathrm{I}_n)\quad P_{n+1}(n+1,x)=(N+D')P_n(n+1,x)+D\frac{dP_n(n+1,x)}{dx}$$

We single out the case $k=n$ because of the fact that this case parallels most closely the Charlier or Hermite polynomial case For in this latter case the n'th derivative of the generating function e^{-x^2} is the product of the generating function and a polynomial of degree n *So in the case of any solution* y *of a Pearson differential equation, the* n *th derivative of* $D^n y$ *is the product of the generating function* y *and a polynomial of degree at most* n

By means of relation (I), we can write down the successive polynomials $P_1(x)$, $P_2(x)$, . The first five polynomials may be written as follows

$$P_1(x) = N,$$

$$P_2(x) = (N-D')P_1(x) + D\frac{dP_1(x)}{dx} = N^2 - ND' + N'D,$$

$$P_3(x) = (N-2D')P_2(x) + D\frac{dP_2(x)}{dx}$$

$$= N^3 - 3N^2D' + 3NN'D + 2ND'^2 - 2N'D'D - NDD'',$$

$$P_4(x) = (N-3D')P_3(x) + D\frac{dP_3(x)}{dx}$$

$$= N^4 - 6N^3D' + 6N^2N'D + 11N^2D'^2 - 14NN'DD' - 4N^2DD'',$$

$$= -6ND'^3 + 6N'DD'^2 + 6NDD'D'' + 3N'^2D^2 - 3N'D^2D'',$$

$$P_5(x) = (N-4D')P_4(x) + D\frac{dP_4(x)}{dx}$$

$$= N^5 - 10N^4D' + 10N^3N'D + 35N^3D'^2 - 50N^2N'DD'$$

$$- 10N^3DD'' - 50N^2D'^3 + 70NN'DD'^2 - 40N^2DD'D''$$

$$+ 15NN'^2D^2 - 25NN'D^2D'' + 24ND'^4 - 24N'DD'^3$$

$$- 36NDD'^2D'' - 20N'^2D^2D' + 24N'D^2D'D'' + 6ND^2D''^2$$

4 Following the analogy with Hermite polynomials, we obtain next a recurrence relation involving the ($n+1$) th, n'th and ($n-1$) th polynomials

Starting with the original differential equation

$$D\frac{dy}{dx} = Ny$$

we take the n th derivative of both sides, which by Leibnitz's theorem on the derivative of a product gives us, since $\frac{d^3 D}{dx^3} = 0$,

$$D\frac{d^{n+1}y}{dx^{n+1}} + nD'\frac{d^n y}{dx^n} + \frac{n(n-1)}{2!}D''\frac{d^{n-1}y}{dx^{n-1}} = N\frac{d^n y}{dx^n} + nN'\frac{d^{n-1}y}{dx^{n-1}}$$

Multiplying this last expression by D^n and collecting terms, we get.

$$D^{n+1}\frac{d^{n+1}y}{dx^{n+1}} + D^n(nD'-N)\frac{d^n y}{dx^n} + D^n\left[\frac{n(n-1)}{2!}D'' - nN'\right]\frac{d^{n-1}y}{dx^{n-1}} = 0$$

Replacing now $D^n\frac{d^n y}{dx^n}$ by $P_n(x)\,y$ and dividing through by y, we get the recurrence relation

$$\text{(II)}\quad P_{n+1}(x) + (nD'-N)P_n(x) + n\left[\frac{(n-1)}{2!}D'' - N'\right]D\,P_{n-1}(x) = 0$$

We note that the coefficients of $P_{n+1}(x)$ and $P_n(x)$ are the same as in relation (I) which we found to be

$$P_{n+1}(x) + P_n(x)(nD'-N) = D\,\frac{dP_n(x)}{dx}$$

Hence

$$\text{(III)}\qquad \frac{dP_n(x)}{dx} = n\left[N' - \frac{(n-1)}{2}D''\right]P_{n-1}(x)$$

or replacing n by $n+1$ we write

$$\frac{d\,P_{n+1}(x)}{dx} = (n+1)\left(N' - \frac{n}{2}D''\right)P_n(x) = (n+1)(a_1 - n\,b_2)\;P_n(x)$$

This equation is the generalized form of the one for Hermite polynomials, viz.

$$\frac{d\,H_n(x)}{dx} = 2\,n\,H_{n-1}(x)$$

5 Relations (I) and (III) may now be used to obtain a second order differential equation. Differentiating (I), we get:

$$P'_{n+1}(x) + (nD'' - N')\,P_n(x) + (nD' - N)\,P_n'(x)$$
$$- D'P_n'(x) - DP_n''(x) = 0$$

Substitution of the value $d\,P_{n+1}(x)/dx$ from (III) gives us

$$\text{(IV)}\qquad \begin{aligned} &DP_n''(x) + \left[N - (n-1)D'\right]P_n'(x) \\ &- n\left[N' - \frac{(n-1)D''}{2}\right]P_n(x) = 0 \end{aligned}$$

We readily see that the relation found for the Hermite polynomials

$$H_n''(x) - 2x\,H_n'(x) + 2\,n\,H_n(x) = 0$$

is a special case of (IV)

Using the lemma previously proved and replacing N by $N + kD'$ we can write (IV) for the polynomials $P_n(k, x)$ and $P_n(n, x)$

$$\text{(IV}_k\text{)}\qquad \begin{aligned} &DP_n''(k, x) + \left[N - (n-k-1)D'\right]P_n'(k, x) \\ &- n\left[N' - \frac{(n-2k-1)}{2}D''\right]P_n(k, x) = 0, \end{aligned}$$

$$(\mathrm{IV}_n)^1 \qquad \begin{aligned} &DP_n''(n,x)+(N+D')P_n'(n,x)\\ &-n\left[N'+\frac{(n+1)}{2}D''\right]P_n(n,x)=0\end{aligned}$$

We recognize the second order differential equations mentioned earlier in this chapter for the polynomials of the Pearson Type

[1]Since D is any expression of the second degree and N is any expression of the first degree, it is obvious that $P_n(x)$ satisfies a linear equation of the second order of the form

$$(A_0+A_1x+A_2x^2)y''+(B_0+B_1x)y'+Cy=0$$

where $C=-n\left[(n-1)A_2+B_1\right]$ It may be shown that if a differential equation of the form considered has as one solution a polynomial of degree n then C must be of the form specified For suppose $Q_n(x)$ satisfies the above differential equation for y Taking the n'th derivative of this equation we get

$$\frac{n(n-1)}{2!}\,2\,A_2(n!a_0)+n\,B_1(n!a_0)+C(n!a_0)=0$$

and solving for C that

$$C=-n\left[(n-1)A_2+B_1\right]$$

It follows from our work that if a differential equation has the form

$$\begin{aligned}&(A_0+A_1x+A_2x^2)y''+(B_0+B_1x)y'\\ &-n\left[(n-1)A_2+B_1\right]y=0\end{aligned}$$

then one solution of this differential equation is a polynomial of degree at most n obtained by finding the solution y of the Pearson differential equation

$$\frac{dy}{dx}=\frac{B_0+B_1x-(A_1+2A_2x)}{A_0+A_1x+A_2x^2}\,y$$

and determining the polynomial

$$P_n(n,x)=\frac{1}{y}\frac{d^n}{dx^n}\left\{\left[A_0+A_1x+A_2x^2\right]^n y\right\}$$

IV, V and VI as well as the Jacobi and Laguerre polynomials as special cases of formula (IV_n) Some further illustrations of (IV_n) are the Tschebycheff[1] and Legendre[2] polynomials The Tschebycheff polynomials are developed from the differential equation

$$\frac{dy}{dx} = \frac{x}{1-x^2}\, y$$

and in this case formula (IV_n) becomes

$$(1-x^2)P_n''(n\,x)-x\,P_n'(n\,x)+n^2P_n(n,x)=0$$

The Legendre polynomials

$$P_n(x) = \frac{1}{2^n n!} \quad \frac{d^n(x^2-1)^n}{dx^n}$$

have as a corresponding differential equation

$$\frac{dy}{dx} = \frac{O\,y}{x^2-1}$$

and in turn formula (IV_n) is written

$$(x^2-1)P_n''(n\,x)+2\,x\,P_n'(n,x) - n(n+1)\,P_n(n,x)=0$$

6 Just as in formula (II) we established a recurrence relation for the polynomials $P_n(x)$, let us now obtain one for the polynomials $P_n(n,x)$

Consider once more the first derivative of $D^k y$, i e

$$\frac{d}{dx}(D^{k+1}y) = (K+1)D'\,D^k y + D^{k+1}y'$$

$$= \left[N+(K+1)D'\right]D^k y$$

[1]R Courant and D Hilbert, op cit, pp 73-74

[2]Ibid, pp 66-69

Taking the n th derivative of both sides of the equation we get

$$\frac{d^{n+1}}{dx^{n+1}}(D^{k+1}y) = \left[N+(k+1)D'\right]\frac{d^n}{dx^n}D^k y$$
$$+\,n\left[N'+(k+1)D''\right]\frac{d^{n-1}}{dx^{n-1}}D^k y$$

Multiplying both sides of the equation by D^{n-k} and replacing $D^{n-k}\frac{d^n}{dx^n}D^k y$ by $P_n(k,x)y$, we have

$$(\mathrm{V}_k)\qquad P_{n+1}(k+1,x) = \left[N+(k+1)D'\right]P_n(k,x) + n\left[N'+(k+1)D''\right]D\,P_{n-1}(k,x)$$

In case we set $k = n$, we may write

$$(\mathrm{V}_n)\qquad P_{n+1}(n+1,x) = \left[N+(n+1)D'\right]P_n(n,x) + n\left[N'+(n+1)D''\right]D\,P_{n-1}(n,x),$$

a recurrence relation similar to (II) and involving the polynomials $P_{n+1}(n+1,x)$, $P_n(n,x)$ and $P_{n-1}(n,x)$

7 Formula (V_n) may be written in still another form corresponding to formula (I), i e a relation consisting of the same terms as (V_n) except that the $(n-1)$ th polynomial $P_{n-1}(n,x)$ is replaced by the first derivative of the n th polynomial $P_n(n,x)$

In order to obtain this relation we return to formula (III),

$$\frac{dP_n(x)}{dx} = n\left[N'-\frac{(n-1)}{2}D''\right]P_{n-1}(x)$$

and substitute for N the value $N+kD'$ and obtain

$$(\mathrm{III}_n)\qquad \frac{dP_n(n,x)}{dx} = n\left[N'+\frac{(n+1)}{2}D''\right]P_{n-1}(n,x)$$

or

$$P_{n-1}(n,x) = \frac{1}{n\left[N' + \left(\frac{n+1}{2}\right)D''\right]} \frac{dP_n(n,x)}{dx}$$

Substituting the value for $P_{n-1}(n,x)$ we thus obtain

$$\text{(VI)} \quad \begin{aligned} P_{n+1}(n+1,x) &= \left[N+(n+1)D'\right]P_n(n,x) \\ &\quad + \frac{N'+(n+1)D''}{N'+\left(\frac{n+1}{2}\right)D''} D \frac{P_n(n,x)}{dx} \end{aligned}$$

From symmetry we might expect the fractional coefficient of the derivative $P_n'(n,x)$ to be unity, but unfortunately this is not the case

8 In looking over the relations existing for the Laguerre polynomials we find one consisting of the first derivatives of the n th and $(n-1)$ th polynomials, and the $(n-1)$ th polynomial,[1] i e

$$P_n'(n,x) - nP_{n-1}'(n,x) = -n P_{n-1}(n-1,x)$$

This relation is a special case of another form of formula (VI) which we obtain in the following manner

Differentiation of (VI) gives us

$$\begin{aligned} \frac{dP_{n+1}(n+1,x)}{dx} &= \left[N'+(n+1)D''\right]P_n(n,x) + \left[N+(n+1)D'\right]\frac{dP_n(n,x)}{dx} \\ &\quad + \frac{N'+(n+1)D''}{N'+\frac{(n+1)}{2}D''} D' \frac{dP_n(n,x)}{dx} + \frac{N'+(n+1)D'}{N'+\frac{(n+1)}{2}D''} D \frac{d^2P_n(n,x)}{dx^2} \end{aligned}$$

Substituting the value for $d^2P_n(n,x)/dx^2$ found in (V) changes this last expression to the form

---

[1] R Courant and D Hilbert, op cit, pp. 77-79

$$\frac{dP_{n+1}(n+1\,x)}{dx} = \left[N'+(n+1)D''\right] P_n(n,x) + \left[N+(n+1)D'\right] \frac{dP_n(n\,x)}{dx}$$

$$+ \frac{N'+(n+1)D''}{N'+\frac{(n+1)}{2}D''} D' \frac{dP_n(n,x)}{dx} + \frac{N+(n+1)D''}{N'+\left(\frac{n+1}{2}\right)D''}$$

$$\left[-(N+D')\frac{dP_n(n\,x)}{dx} + n\left(\frac{n+1}{2}D''+N'\right)P_n(n,x)\right],$$

which reduces to

$$\text{(VII)}\quad \begin{aligned} \frac{dP_{n+1}(n+1,x)}{dx} &= (n+1)\left[N'+(n+1)D''\right] P_n(n,x) \\ &+ \left\{\left[N+(n+1)D'\right] - \frac{N'+(n+1)D''}{N'+\frac{(n+1)}{2}D''}\, N\right\} \frac{dP_n(n,x)}{dx} \end{aligned}$$

The special equation mentioned for the Laguerre polynomials will be recognized as a special case of formula (VII) if we recall that for the Laguerre polynomials the differential equation is of the form

$$\frac{dy}{dx} = \frac{p-x}{x}\,y$$

Substitution of x for D and $(p-x)$ for N reduces (VII) to

$$P'_{n+1}(n+1,x) = -(n+1)P_n(n,x) + (n+1)P'_n(n,x)$$

9 In this chapter we have defined two general types of polynomials $\qquad P_n(x) = \frac{D^n}{y}\frac{d^n y}{dx^n}$

and $\qquad P_n(k\,x) = \frac{D^{n\,k}}{y}\frac{d^n}{dx^n} D^k y$

The relationships for these polynomials $P_n(x)$ and $P_n(k,x)$ were derived without using the form of the solution of the differential equation Two fundamental formulas were derived, for $P_n(x)$

$$\text{(I)}\qquad P_{n+1}(x) = (N-nD')P_n(x) + D\frac{dP_n(x)}{dx}$$

and for $P_n(n, x)$ the corresponding formula

$$\text{(VI)}\quad P_{n+1}(n+1, x) = \left[N + (n+1)D'\right] P_n(n, x) + \frac{N' + (n+1)D''}{N' + \frac{(n+1)}{2} D''} D \frac{dP_n(n x)}{dx}$$

Two successive polynomials were shown to be related by the relations, for $P_n(x)$.

$$\text{(III)}\qquad \frac{dP_n(x)}{dx} = n\left[N' - \frac{(n-1)}{2} D''\right] P_{n-1}(x)$$

and for $P_n(n, x)$

$$(\text{III}_n)\qquad \frac{dP_n(n, x)}{dx} = n\left[N' + \frac{(n+1)}{2} D''\right] P_{n-1}(n, x)$$

In addition we found that it was possible to set up recurrence relations involving the $(n+1)$ th, n th and $(n-1)$ th polynomials and found these to be, for $P_n(x)$

$$\text{(II)}\quad P_{n+1}(x) + (nD' - N) P_n(x) + n\left[\frac{n-1}{2!} D'' - N'\right] D\ P_{n-1}(x) = 0$$

and for $P_n(n, x)$

$$(\text{V}_n)\quad P_{n+1}(n+1, x) = \left[N + (n+1)D'\right] P_n(n, x) + n\left[N' + (n+1)D''\right] D P_{n-1}(n x)$$

We further succeeded in developing a second order differential equation for the n'th polynomial $P_n(x)$.

$$\text{(IV)}\ DP_n''(x) + \left[N - (n-1)D'\right] P_n'(x) - n\left[N' - \frac{(n-1)}{2} D''\right] P_n(x) = 0$$

and for $P_n(n, x)$

$$(\text{IV}_n)\ DP_n''(n, x) + (N + D')P_n'(n, x) - n\left[N' + \frac{(n+1)}{2} D''\right] P_n(n, x) = 0$$

We also showed that we could derive a relation between the derivatives of the polynomials $P_{n+1}(n+1, x)$, $P_n(n, x)$ and the polynomial $P_n(n, x)$

$$\text{(VII)}\quad \frac{dP_{n+1}(n+1,x)}{dx} = (n+1)\left[N' + (n+1)D''\right]P_n(n\,x) +$$

$$\left\{\left[N + (n+1)D'\right] - \frac{N' + (n+1)D''}{N' + \left(\frac{n+1}{2}\right)D''}\, N\right\}\frac{dP_n(n\,x)}{dx}.$$

Finally, we noted that all of these formulas and relations apply to the Hermite, Jacobi, Tschebycheff and Legendre polynomials as well as the polynomials derived for the Pearson Type IV, V and VI curves by Romanovsky

CHAPTER III

1 So far the discussion in this paper has been limited to the treatment of the Gram-Charlier series where the constants $A_0, A_1, A_2, \quad , A_n,$ depend upon polynomials in x which are independent of the function $F(x)$, and the generating function $f(x)$ is a solution of the Pearson differential equation, the functions $F(x)$ and $f(x)$ being defined as continuous functions The work in mathematical statistics involves not only the use of the continuous variate and the continuous function but also the case of the discrete variate and the discontinuous function where this function is defined for equally spaced values

In dealing with the continuous variate we make use of the theory of the differential and integral calculus, or the calculus of limits, as it is sometimes called. On the other hand, for the discrete variate we turn to the theory of the calculus of finite differences. Further, it usually happens that there exists a parallelism between results based on the derivative and integral and those based on the finite differences and summations As a consequence, it seems natural to attempt to derive results for the finite difference case paralleling those contained in the first half of this paper The second part of this paper is devoted to this purpose The first of the two following chapters considers matters pertaining to Charlier's Type B series which is the finite difference parallel to the Type A series, while the next chapter is devoted to the polynomials connected with the finite difference parallel of the Pearson differential equation

Charlier in the second half of his article[1] "Ueber die Dar-

*C V L Charlier, op cit, pp 23-35

stellung willkurlicher Funktionen" considers a real valued function $F(x)$ and asserts that it may be formally expanded in terms of another function and its successive differences Stated as a theorem, this may be written as follows·

CHARLIER'S THEOREM FOR SERIES B· *Any real valued function $F(x)$ which vanishes for $x = \infty$ and $-\infty$, may be formally expanded in terms of another function $g(x)$ and its successive differences in the form*

$$(\mathrm{B})\ F(x) = B_0 g(x) + B_1 \Delta g(x) + B_2 \Delta^2 g(x) + \quad + B_n \Delta^n g(x) +$$

where $g(x)$ possesses the properties

(a) $g(x)$ *and its differences are defined for all real values of* x,

(b) $g(x)$ *and its differences vanish for* $x = +\infty$ *and* $-\infty$,

(c) $x^m \Delta^n g(x) \Big|_{-\infty}^{+\infty} = 0$ *for all real values of* m *and* n

(d) $\Delta^{-1} g(x) \Big|_{-\infty}^{+\infty} \neq 0$

Paralleling the theory of the first half of his paper, Charlier determines the constants $B_0, B_1, B_2, \quad , B_n,$ and finds that they may be expressed by the equation

$$B_n = \sum_{-\infty}^{+\infty} Q_n(x) F(x) = \Delta^{-1} Q_n(x) F(x) \Big|_{\infty}^{+\infty}$$

where $Q_n(x)$ is a polynomial in x of degree not greater than n Analyzing the answers that he obtains for $Q_n(x)$, we find that these polynomials form a uniquely determined set of polynomials $Q_0(x)$, $Q_1(x)$, $Q_2(x)$, $\quad$, $Q_n(x)$, $\quad$, $Q_n(x)$ at most of degree n, biorthogonal in the sum sense to the successive differences of the function $g(x)$, i e they satisfy the biorthogonality conditions for the inverse of differences

$$\Delta^{-1} Q_n(x)\,\Delta^m g(x)\Big|_{-\infty}^{+\infty} \begin{matrix} =0 & \text{for} & n \neq m \\ 1 & \text{for} & n = m. \end{matrix}$$

Charlier does not observe that the polynomials $Q_n(x)$ bear a definite relation to one another, i. e.

$$\Delta Q_n(x) = -Q_{n-1}(x+1),$$

a relation similar to the one found for the polynomials $P_n(x)$ in Chapter I. We may state these facts in the following theorem:

THEOREM: *If $g(x)$ satisfy the conditions* (a), (b), (c), *and* (d) *of Charlier's Theorem for series* B *and if $Q_o(x)$, $Q_1(x)$, . . ., $Q_n(x)$, . . . is the system of polynomials in x, $Q_n(x)$ of degree at most n, which is biorthogonal to $f(x)$ and its differences, i e. satisfies the conditions*

$$\Delta^{-1} Q_n(x)\,\Delta^m g(x)\Big|_{-\infty}^{+\infty} \begin{matrix} = 0 \text{ for } n \neq m \\ = 1 \text{ for } n = m \end{matrix}$$

then

$$\Delta Q_n(x) = -Q_{n-1}(x+1).$$

The proof requires the use of the finite integration by parts formula:

$$\Delta^{-1} u_x v_x = u_x \Delta^{-1} v_x - \Delta^{-1}\left[\Delta u_x \cdot \Delta^{-1} v_{x+1}\right].$$

Applying this formula we get

$$\Delta^{-1} Q_n(x)\Delta^m g(x)\Big|_{-\infty}^{+\infty} = Q_n(x)\,\Delta^{m-1} g(x)\Big|_{-\infty}^{+\infty}$$

$$-\Delta^{-1}\Big[\Delta Q_n(x)\,\Delta^{m-1} g(x+1)\Big]_{-\infty}^{+\infty}$$

The first term on the right hand side vanishes due to condition (c) of the theorem of Charlier. Comparing the term which

remains, i e.

$$-\Delta^{-1}\left[\Delta Q_n(x)\Delta^{m-1}g(x+1)\right]_{-\infty}^{+\infty} \begin{matrix} = 0 & \text{for } n \neq m \\ = 1 & \text{for } n = m \end{matrix}$$

with the biorthogonality condition

$$\Delta^{-1}\left[Q_{n-1}(x+1)\Delta^{m-1}g(x+1)\right]_{-\infty}^{+\infty} \begin{matrix} = 0 & \text{for } n \neq m \\ = 1 & \text{for } n = m \end{matrix}$$

we conclude that

$$\Delta Q_n(x) = -Q_{n-1}(x+1)$$

This theorem enables us to find the terms of the n th polynomial by taking the negative of the integral of the $(n-1)$ th polynomial, except for the constant of integration. Following the suggestion in our first chapter, we may also determine this constant We have

$$Q_n(x) = -\Delta^{-1}Q_{n-1}(x+1)\Big|_0^x + C$$

and the simple biorthogonality condition

$$\Delta^{-1}Q_n(x)g(x)\Big|_{-\infty}^{+\infty} = 0.$$

It follows that

$$\Delta^{-1}\left[-\Delta^{-1}Q_{n-1}(x+1)+C\right]_0^x g(x)\Big|_{-\infty}^{+\infty} = 0$$

and solving for C we get

$$C = \frac{\Delta^{-1}\left[\Delta^{-1}Q_{n-1}(x+1)\right]_0^x g(x)\Big|_{-\infty}^{+\infty}}{\Delta^{-1}g(x)\Big|_{-\infty}^{+\infty}}$$

We may therefore determine the polynomials $Q_n(x)$ from the polynomials next preceding by the formula

$$Q_n(x) = -\Delta^{-1}Q_{n-1}(x+1)\Big|_0^x + \frac{\Delta^{-1}\left[\Delta^{-1}Q_{n-1}(x+1)\right]_0^x g(x)\Big|_{-\infty}^{+\infty}}{\Delta^{-1}g(x)\Big|_{-\infty}^{+\infty}}$$

If we adopt the Charlier notation

$$\mathcal{E}_m = \sum_{-\infty}^{+\infty} x^m g(x) = \Delta^{-1} x^m g(x) \Big|_{-\infty}^{+\infty}$$

and the common notation $x^{(m)} = x(x-1)(x-2) \quad \cdot \quad (x-m+1)$ and observe that $Q_0(x) = 1/\mathcal{E}_0$ and that

$$\Delta^{-1} x^{(m)} = \frac{x^{(m+1)}}{m+1}$$

we may obtain the polynomials $Q_1(x)$, $Q_2(x)$, ... without much computation as follows:

$$Q_1(x) = -\Delta^{-1} Q_0(x+1) \Big|_0^x + \frac{\Delta^{-1}\left[\Delta^{-1} Q_0(x+1)\right]_0^x g(x) \Big|_{-\infty}^{+\infty}}{\Delta^{-1} g(x) \Big|_{-\infty}^{+\infty}}$$

$$= -\frac{x}{\mathcal{E}_0} + \frac{\mathcal{E}_1}{\mathcal{E}_0^2}$$

$$Q_2(x) = -\Delta^{-1} Q_1(x+1) \,{}_0^x + \frac{\Delta^{-1}\left[\Delta^{-1} Q_1(x+1)\right]_0^x g(x) \Big|_{-\infty}^{+\infty}}{\Delta^{-1} g(x) \Big|_{-\infty}^{+\infty}}$$

$$= \frac{(x+1)^{(2)}}{\underline{|2}\, \mathcal{E}_0} - \frac{\mathcal{E}_1(x+1)}{\mathcal{E}_0^2} + \frac{2\mathcal{E}_1^2 + \mathcal{E}_1\mathcal{E}_0 - \mathcal{E}_2\mathcal{E}_0}{\underline{|2}\, \mathcal{E}_0^3}$$

$$\text{or } \underline{|2}\, \mathcal{E}_0^3 Q_2(x) = \mathcal{E}_0^2 x^2 - \mathcal{E}_0 x(2\mathcal{E}_1 - \mathcal{E}_0) + 2\mathcal{E}_1^2 - \mathcal{E}_2\mathcal{E}_0 - \mathcal{E}_1\mathcal{E}_0$$

$$Q_3(x) = -\Delta^{-1} Q_2(x+1) \Big|_0^x + \frac{\Delta^{-1}\left[\Delta^{-1} Q_2(x+1)\right]_0^x g(x) \Big|_{-\infty}^{+\infty}}{\Delta^{-1} g(x) \Big|_{-\infty}^{\infty}}$$

$$= -\frac{(x+2)^{(3)}}{\underline{|3}\, \mathcal{E}_0} + \frac{\mathcal{E}_1(x+2)^{(2)}}{\underline{|2}\, \mathcal{E}_0^2} - \frac{(2\mathcal{E}_1^2 + \mathcal{E}_1\mathcal{E}_0 - \mathcal{E}_2\mathcal{E}_0)(x+2)}{\underline{|2}\, \mathcal{E}_0^3}$$

$$+ \frac{\mathcal{E}_3\mathcal{E}_0^2 - 3\mathcal{E}_2\mathcal{E}_0^2 - 6\,\mathcal{E}_2\mathcal{E}_1\mathcal{E}_0 + 6\mathcal{E}_1^3 + 6\mathcal{E}_1^2\mathcal{E}_0 + 2\mathcal{E}_1\mathcal{E}_0^2}{\underline{|3}\, \mathcal{E}_0^4}$$

$$\text{or } \underline{|3}\, \mathcal{E}_0^4 Q_3(x) = -\mathcal{E}_0^3 x^3 + 3\mathcal{E}_0^2 x^2(\mathcal{E}_1 - \mathcal{E}_0)$$

$$- \mathcal{E}_0 x(2\mathcal{E}_0^2 - 6\mathcal{E}_1\mathcal{E}_0 - 3\mathcal{E}_2\mathcal{E}_0 + 6\mathcal{E}_1^2) + \mathcal{E}_0^2\mathcal{E}_3 + 3\mathcal{E}_2\mathcal{E}_0^2 + 2\mathcal{E}_1\mathcal{E}_0^2$$

$$- 6\mathcal{E}_2\mathcal{E}_1\mathcal{E}_0 - 6\mathcal{E}_1^2\mathcal{E}_0 + 6\mathcal{E}_1^3$$

. . . .

These results differ slightly from those obtained by Charlier in his article. This is due to the definition for differences used by Charlier, viz.:

$$\Delta g(x) = g(x) - g(x-1)$$

while we have used the definition

$$\Delta g(x) = g(x+1) - g(x) \ .$$

Denoting the difference

$$g(x) - g(x-1) \text{ by } \delta g(x)$$

Charlier determines a set of polynomials $T_n(x)$ satisfying the conditions ,

$$\delta^{-1}\left[T_n(x)\,\delta^m g(x)\right] \begin{array}{ll} = 0 & \text{for } m \neq n \\ = 1 & \text{for } m = n \end{array}$$

As a consequence by paralleling the reasoning above one proves easily that the $T_n(x)$ satisfy the recurrence relation

$$T_n(x+1) - T_n(x) = -T_{n-1}(x) \ .$$

By using this relation and the fact that

$$\delta^n g(x+n) = \Delta^n g(x)$$

it can be shown without much difficulty that

$$T_n(x+n-1) = Q_n(x)$$

The theorem proved in Ch. 1, par. 2, could no doubt be paralleled by using finite difference theory Since the method of procedure is obvious there seems to be no need of taking it up in detail.

We have succeeded in showing in this chapter that the problem of determining the constants for the Charlier Type B series closely parallels the work of the first chapter and that these constants are readily obtained by using the biorthogonality conditions for finite differences

CHAPTER IV

Polynomials Connected with the Pearson Difference Equation

1 In Chapter II we referred to certain solutions $f(x)$ of the Pearson differential equation and noted that graphically, these functions represented types of curves used in statistical work Paralleling this work, we would expect to find that a difference equation similar in composition to the Pearson differential equation would have as solutions functions $g(x)$ which could be used to represent data consisting of discrete variates

Carver, in an article in the "Handbook of Mathematical Statistics,"* suggests the use of a difference equation corresponding to the Pearson differential equation, i. e.

$$\Delta u_x = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots} u_x ,$$

a difference equation with a numerator of the first and denominator of any desired degree in x. If we confine our work to a denominator of degree at most of the second in x, we should be able to obtain results comparing very favorably with those obtained in the second chapter

An illustration of a solution of this difference equation found in Charlier's article "Ueber die Darstellung willkürlicher Funktionen,"[2] is the well known Poisson exponential function

$$\psi(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

[1]H C. Carver, "Frequency Curves," Handbook of Mathematical Statistics (H. L Rietz, Editor), Chapter VII, pp. 111-114.

[2]C V. L. Charlier, op cit p. 33

This function satisfies the difference equation

$$\Delta u_x = \frac{\lambda - x - 1}{x+1} \; u_x$$

and this equation is recognized as a special form of the Pearson difference equation. If we take the successive differences of this Poisson exponential function, we find that these give rise to a unique set of polynomials. These polynomials may be written in the following form:

$$Q_1(x) = \lambda - (x+1) \,,$$

$$Q_2(x) = \lambda^2 - 2\lambda(x+2) + (x+2)(x+1) \,,$$

or making use of the usual difference notation for

$x^{(m)} = x(x-1)(x-2) \cdot \quad (x-m+1)$, we write

$$Q_2(x) = \lambda^2 - 2\lambda(x+2) + (x+2)^{(2)} ,$$

$$Q_3(x) = \lambda^3 - 3\lambda^2(x+3) + 3\lambda(x+3)^{(2)} - (x+3)^{(3)} ,$$

or $\quad Q_3(x-3) = \lambda^3 - 3\lambda^2 x + 3\lambda x^{(2)} - x^{(3)} ,$

. . . .

$$Q_n(x) = \lambda^n - {}_nC_1\lambda^{n-1}(x+n) + {}_nC_2\lambda^{n-2}(x+n)^{(2)} + \cdots + (-1)^n (x+n)^{(n)}$$

or $Q_n(x-n) = \lambda^n - {}_nC_1\lambda^{n-1}x + {}_nC_2\lambda^{n-2}x^{(2)} + \quad + (-1)^n x^{(n)} .$

These polynomials have the same form as that for the binomial expansion $(\lambda - x)^n$, particularly if we use the difference notation for representing powers of x In other words, we might look upon the n th polynomial as being defined as

$$Q_n(x-n) = \left[\lambda - x^{(1)}\right]^n$$

A careful examination of these polynomials brings out the fact that consecutive ones are related to each other, viz., that we have,

$$\Delta Q_n(x) = -nQ_{n-1}(x+1)$$

This relation is similar to the one found for Hermite polynomials.

The fact that the Charlier Type A series in Chapter II consisted of successive derivatives and that the derivatives of the solutions of the Pearson differential equation led to a system of polynomials definitely related to one another, gave rise to the theory developed in that chapter We found that it was not necessary in this theory to consider the form of the solution of the equation, but that a set of general polynomials could be set up which satisfied all the properties of the special polynomials. The Charlier Type B series consists of successive differences of a function $g(x)$ and it is quite natural for us to suspect that we can develop for the solutions of the Pearson difference equation a corresponding theory on polynomials.

This question of obtaining a system of polynomials from the solutions of the Pearson difference equation

$$(1) \qquad \Delta u_x = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} u_x ,$$

numerator of the first degree and denominator of the second degree, will concern us in this chapter. We shall further show that these polynomials are related to one another by means of first and second order difference relations and by means of recurrence relations involving the $(n+1)$ th, n th and $(n-1)$ th polynomials, and shall illustrate these equations with the Poisson exponential function.

2. For convenience denote the numerator $(a_0 + a_1 x)$ in equation (1) by N_x and the denominator $(b_0 + b_1 x + b_2 x^2)$ by D_x. We may then define a set of polynomials by the following theorem:

THEOREM: *If* u_x *is a non-identically zero solution of*

$$\Delta u_x = \frac{N_x}{D_x} u_x$$

then $\frac{1}{u_x} D_x D_{x+1} D_{x+2} \ D_{x+n-1} \Delta^n u_x$ *is a polynomial of degree at most* n, i. e. $Q_n(x)$.

The proof will proceed by mathematical induction. If we recall the formula for the difference of a product

$$\Delta [u_x v_x] = v_x \Delta u_x + u_{x+1} \Delta v_x = v_x \Delta u_x + (u_x + \Delta u_x) \Delta v_x ,$$

we obtain by differencing

$$D_x \Delta u_x = N_x u_x = Q_1(x) u_x$$

the equation

$$D_{x+1} \Delta^2 u_x + \Delta u_x \Delta D_x = [Q_1(x) + \Delta Q_1(x)] \Delta u_x + u_x \cdot \Delta Q_1(x)$$

Using the value for Δu_x from the original difference equation and multiplying the equation through by D_x, we obtain:

$$D_x D_{x+1} \Delta^2 u_x = [N_x Q_1(x) + D_x \Delta Q_1(x) + N_x \Delta Q_1(x) + N_x \Delta D_x] u_x$$

Since the coefficient of u_x is a polynomial of degree at most 2 in x, we write

$$D_x D_{x+1} \Delta^2 u_x = Q_2(x) u_x$$

Let us now assume that the statement holds for $m \leq n$, i. e.

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n-1} \Delta^n u_x = Q_n(x) u_x .$$

Differencing both sides of this equation gives us

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n-1} \Delta^{n+1} u_x + (\Delta^n u_x + \Delta^{n+1} u_x)(\Delta D_x D_{x+1} D_{x+2} \cdots D_{x+n-1})$$

$$= Q_n(x) u_x + (u_x + \Delta u_x) \Delta Q_n(x).$$

Now

$$\Delta D_x D_{x+1} \cdots D_{x+n-1} = D_{x+1} D_{x+2} \cdots D_{x+n} - D_x D_{x+1} \cdots D_{x+n-1} =$$

$$= (D_{x+1} \cdots D_{x+n-1})(D_{x+n} - D_x).$$

Hence by the definition of $Q_n(x)$

$$\Delta \left[D_x D_{x+1} \cdots D_{x+n-1} \Delta^n u_x \right] = \frac{D_{x+n} - D_x}{D_x} Q_n(x) u_x$$

Substituting these values in the above equation as well as the value for Δu_x from (1) and multiplying by D_x, the equation reduces to

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n} \Delta^{n+1} u_x = \Big[N_x Q_n(x) + D_x \Delta Q_n(x)$$

$$+ N_x \Delta Q_n(x) - D_{x+n} Q_n(x) + D_x Q_n(x) \Big] u_x$$

The coefficient of u_x on the right hand side is a polynomial of degree at most n in x. We therefore conclude that

$$D_x D_{x+1} \quad D_{x+n} \Delta^{n+1} u_x = Q_{n+1}(x) u_x$$

We have also succeded in deriving a relation similar to relation (I) of Chapter II, i. e.

$$\text{(XI)} \qquad \begin{aligned} Q_{n+1}(x) &= (N_x + D_x - D_{x+n}) Q_n(x) \\ &\quad + (N_x + D_x) \Delta Q_n(x), \end{aligned}$$

a relation which shows that the ($n+1$) th polynomial is made up of the n th polynomial and the difference of the n th polynomial. This relation differs from relation (I) in the fact that the coefficient of $\Delta Q_n(x)$ is $N_x + D_x$ instead of D_x . This change seems to be connected with the fact that the original difference equation

$$D_x \Delta u_x = N_x u_x$$

can also be written

$$u_{x+1} = \frac{N_x + D_x}{D_x} u_x$$

Formula (XI) may also be written

$$\text{(XI)} \quad Q_{n+1}(x) = (N_x + D_x) Q_n(x+1) - D_{x+n} \; Q_n(x)$$

since $\quad Q_n(x) + \Delta Q_n(x) = Q_{n+1}(x)$.

It seems advisable to adopt a notation for the term

$$D_x D_{x+1} D_{x+2} \cdots D_{x+n-1}$$

since it will continue to be involved in the work that is to follow. The difference notation $x^{(m)} = x(x-1)(x-2) \cdots (x-m+1)$ suggests that we use the symbol $D_x^{(n)}$, i. e.

$$D_x^{(n)} = D_x D_{x-1} D_{x-2} \qquad D_{x-n+1}$$

Then we will have

$$D_x D_{x+1} D_{x+2} \cdot \quad D_{x+n-1} = D_{x+n-1}^{(n)}$$

and

$$\Delta D_{x+n-1}^{(n)} = D_{x+n} D_{x+n-1} \cdots D_{x+2} D_{x+1}$$
$$- D_{x+n-1} D_{x+n-2} \quad D_{x+1} D_x = (D_{x+n} - D_x) D_{x+n-1}^{(n-1)}$$

3. We may also define the general polynomials $Q_n(m, x)$ where m is any integer, by means of a theorem as follows:

THEOREM: *If u_x is a non-identically zero solution of the difference equation* (1), *then*

$$\frac{D_{x-m+n-1}^{(n)} \; \Delta^n \left[D_{x-1}^{(m)} \; u_x\right]}{D_{x-1}^{(m)} u_x}$$

is a polynomial $Q_n(m, x)$, and $Q_n(m, x)$ is at most of degree n in x. In particular if $m = n$, we have

$$\frac{1}{u_x} \Delta^n \left[D_{x-1}^{(n)} \; u_x\right]$$

is a polynomial in x of degree at most n.

This theorem may be proved by using the following lemma:

LEMMA: *If u_x satisfy the difference equation* (1), *then $D_{x-1}^{(m)} u_x$, where m is any positive integer, satisfies a difference equation of the same type, viz.*

$$\Delta\left[D_{x-1}^{(m)} u_x\right] = \frac{D_{x-1}^{(m)} u_x \left[N_x + D_x - D_{x-m}\right]}{n_{x-m}} .$$

The proof proceeds easily by mathematical induction.
For $m = 1$ we have

$$\Delta\left[D_{x-1}u_x\right]=D_x\Delta u_x+u_x\Delta D_{x-1}$$

$$=N_x u_x+\Delta D_{x-1}u_x$$

$$=D_{x-1}u_x\left[\frac{N_x+D_x-D_{x-1}}{D_{x-1}}\right],$$

For $m=2$, we get

$$\Delta[D_{x-2}D_{x-1}u_x]=D_{x-1}\Delta[D_{x-1}u_x]+D_{x-1}u_x\Delta D_{x-2}$$

$$=D_{x-2}D_{x-1}u_x\left[\frac{N_x+\Delta D_{x-1}+\Delta D_{x-2}}{D_{x-2}}\right],$$

or $$\Delta[D_{x-1}^{(2)}u_x]=D_{x-1}^{(2)}u_x\left[\frac{N_x+D_x-D_{x-2}}{D_{x-2}}\right]$$

Let us assume that it holds for the mth case, i. e.

$$\Delta[D_{x-1}^{(m)}u_x]=D_{x-1}^{(m)}u_x\left[\frac{N_x+D_x-D_{x-m}}{D_{x-m}}\right].$$

Then

$$\Delta[D_{x-m-1}D_{x-1}^{(m)}u_x]=D_{x-m}\Delta[D_{x-1}^{(m)}u_x]+D_{x-1}^{(m)}u_x\Delta D_{x-m-1}$$

$$=D_{x-1}^{(m)}u_x\left[N_x+D_x-D_{x-m}+D_{x-m}-D_{x-m-1}\right]$$

$$=D_{x-1}^{(m+1)}u_x\left[\frac{N_x+D_x-D_{x-m-1}}{D_{x-m-1}}\right].$$

Making use of this lemma in proving the last theorem, we note that

$$\Delta^2[D_{x-1}^{(m)}u_x]=D_{x-1}^{(m)}u_x\quad\frac{[N_x+D_x-D_{x-m}]}{D_{x-m}D_{x-m+1}}$$

or $$D_{x-m}D_{x-m+1}\Delta^2[D_{x-1}^{(m)}u_x]=D_{x-1}^{(m)}u_x\ [N_x+D_x-D_{x-m}]$$

and in general that

$$\Delta^{n}\left[D_{x-1}^{(m)} u_x\right] = \frac{D_{x-1}^{(m)} u_x \left[Q_n(m, x)\right]}{D_{x-m+n-1}^{(n)}}$$

or $$D_{x-m+n-1}^{(n)} \Delta^{n}\left[D_{x-1}^{(m)} u_x\right] = D_{x-1}^{(m)} u_x \; Q_n(m, x).$$

In particular, if $m = n$, we define the polynomials $Q_n(n, x)$ as $\Delta^{n}\left[D_{x-1}^{(n)} u_x\right] = Q_n(n, x)\, u_x$ which relation is of interest because the Δ^{n} has no D_x as multiplier. Any result derived for the polynomials $Q_n(x) = \frac{1}{u_x} D_{x+n-1}^{(n)} \Delta^{n} u_x$ where u_x is a solution of the difference equation (1) can now be extended to the polynomials $Q_n(m, x) = \dfrac{\Delta^{n}\left[D_{x-1}^{(m)} u_x\right]}{D_{x-1}^{(m-n)} u_x}$ by replacing N_x by $(N_x + D_x - D_{x-m})$ and D_x by D_{x-m}. For example, relation (XI) becomes

$$(\mathrm{XI}_m) \quad \begin{aligned} Q_{n+1}(m+1, x) &= (N_x + D_x - D_{x-m+n-1})\, Q_n(m+1, x) \\ &\quad + (N_x + D_x)\, \Delta Q_n(m+1, x) \end{aligned}$$

and when $m = n$, this relation reduces to

$$(\mathrm{XI}_n) \quad \begin{aligned} Q_{n+1}(n+1, x) &= (N_x + D_x - D_{x-1}) Q_n(n+1, x) + (N_x + D_x) \Delta Q_n(n+1, x) \\ &= (N_x + \Delta D_{x-1}) Q_n(n+1, x) + (N_x + D_x) \Delta Q_n(n+1, x). \end{aligned}$$

4 In analogy with the work of chapter II, we next proceed to find a recurrence relation involving the $(n+1)$ th, n th and $(n-1)$ th of the polynomials $Q(x)$ We take the n th difference of both sides of the equation

$$D_x \Delta u_x = N_x u_x$$

by making use of the formula for the n th difference of a product

$$\Delta^n [u_x v_x] = v_x \Delta^n u_x + n \Delta v_x \Delta^{n-1} u_{x+1}$$

$$+ \frac{n(n-1)}{2!} \Delta^2 v_x \Delta^{n-2} u_{x+2} + \cdots .$$

We then obtain the equation

$$D_x \Delta^{n+1} u_x + n \Delta D_x \Delta^n u_{x+1} + \frac{n(n-1)}{2!} \Delta^2 D_x \Delta^{n-1} u_{x+2} =$$

$$N_x \Delta^n u_x + n \Delta N_x \Delta^{n-1} u_{x+1} ,$$

$\Delta^3 D_x$ and $\Delta^2 N_x$ being equal to zero. Multiplying through by $D_{x+n}^{(n)}$ we get

$$D_{x+n}^{(n)} \Delta^{n+1} u_x + n D_{x+n}^{(n)} \Delta D_x \Delta^n u_{x+1}$$

$$+ \frac{n(n-1)}{2!} D_{x+n}^{(n)} \Delta^2 D_x \Delta^{n-1} u_{x+2}$$

$$= N_x D_{x+n}^{(n)} \Delta^n u_x + n D_{x+n}^{(n)} \Delta N_x \Delta^{n-1} u_{x+1}$$

But $u_{x+1} = u_x + \Delta u_x$ and $u_{x+2} = u_x + 2\Delta u_x + \Delta^2 u_x$.

Substituting these values in the last equation and using the defiinition for the polynomials $Q_n(x)$, we obtain:

$$Q_{n+1}(x) u_x + \frac{n \Delta D_x}{D_x} \left[D_{x+n} Q_n(x) + Q_{n+1}(x) \right] u_x$$

$$+ \frac{n(n-1)}{2!} \frac{D_{x+1}}{D_{x+1}} \frac{\Delta^2 D_x}{D_x} \left[D_{x+n}^{(2)} Q_{n-1}^{(x)} + 2 D_{x+n} Q_n(x) + Q_{n+1}(x) \right] u_x$$

$$= \frac{N_x}{D_x} D_{x+n} Q_n(x) u_x + \frac{n D_{x+n} \Delta N_x}{D_x} \left[D_{x+n-1} Q_{n-1}(x) + Q_n(x) \right] u_x$$

Dividing through by u_x and collecting like terms, this expression reduces to

$$\left[1+\frac{n\Delta D_x}{D_x}+\frac{n(n-1)}{2}\frac{\Delta^2 D_x}{D_x}\right]Q_{n+1}(x)+\left[\frac{nD_{x+n}\Delta D_x}{D_x}+\frac{n(n-1)\Delta^2 D_x D_{x+n}}{D_x}\right.$$

$$\left.-\frac{N_x D_{x+n}}{D_x}-\frac{nD_{x+n}\Delta N_x}{D_x}\right]Q_n(x)+\left[\frac{n(n-1)}{2!D_x}D_{x+n-1}D_{x+n}\Delta^2 D_x\right.$$

$$\left.-\frac{nD_{x+n-1}D_{x+n}\Delta N_x}{D_x}\right]Q_{n-1}(x)=0$$

Now we know that

$$u_{x+n}=u_x+n\Delta u_x+\frac{n(n-1)}{2!}\Delta^2 u_x+\cdots\cdot$$

and so we may write D_{x+n} and N_{x+n} in this same form, i. e.

$$D_{x+n}=D_x+n\Delta D_x+\frac{n(n-1)}{2!}\Delta^2 D_x,$$

$$\Delta D_{x+n}=\Delta D_x+n\Delta^2 D_x,$$

and $$N_{x+n}=N_x+n\Delta N_x$$

the third and higher differences of D_x and the second and higher differences of N_x being equal to zero. The coefficient of $Q_{n+1}(x)$ reduces to $\frac{2D_{x+n}}{D_x}$ and the coefficient of $Q_n(x)$ also reduces to a simpler form. Dividing through by $\frac{2D_{x+n}}{D_x}$ we finally get the recurrence relation:

$$\text{(XII)}\qquad \begin{aligned} &Q_{n+1}(x)+(n\Delta D_{x+n-1}-N_{x+n})Q_n(x)\\ &+nD_{x+n-1}\left[\frac{(n-1)}{2}\Delta^2 D_x-\Delta N_x\right]Q_{n-1}(x)=0\end{aligned}$$

i. e. the $(n+1)$ th polynomial may be obtained from the n th and $(n-1)$ th polynomials.

In Chapter II we found that relations (I) and (II) were identical for the first two terms, and as a consequence we equated the third terms and obtained a relation between the derivative of

a polynomial $P_n(x)$ and the polynomial preceding it. In order that we may obtain a similar expression for the difference polynomials, we must change the appearance of formula (XII).

By lowering the degree in formula (XI) from n to $n-1$ and solving for $D_{x+n-1}\, Q_{n-1}(x)$ we find that

$$D_{x+n-1}\, Q_{n-1}(x) = (N_x + D_x)\, Q_{n-1}(x+1) - Q_n(x).$$

Substitution of this relation in formula XII gives

$$Q_{n+1}(x) = (N_{x+n} - n\Delta D_{x+n-1})\, Q_n(x)$$

$$+ n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right]\left[(N_x + D_x)\, Q_{n-1}(x+1) - Q_n(x)\right]$$

or $Q_{n+1}(x) = \left[N_{x+n} - n\Delta N_x - n\Delta D_{x+n-1} + \frac{n(n-1)}{2}\Delta^2 D_x\right] Q_n(x)$

$$+ n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right](N_x + D_x)\, Q_{n-1}(x+1).$$

Just as in Chapter IV, paragraph 3, the coefficient of $Q_n(x)$ reduces and becomes the same as the coefficient of $Q_n(x)$ in formula (XI) and we have

$$Q_{n+1}(x) = (N_x + D_x - D_{x+n})\, Q_n(x)$$

(XII′)

$$+ n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right](N_x + D_x)\, Q_{n-1}(x+1).$$

We therefore conclude that

(XIII) $\Delta Q_n(x) = n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right] Q_{n-1}(x+1),$

a relation expressing the difference of a polynomial $Q_n(x)$ in terms of the next preceding polynomial in $(x+1)$, i. e.

$Q_{n-1}(x+1)$. For the polynomial $Q_n(n, x)$, formula (XII) may be written in the form

$$(\mathrm{XIII}_n) \quad \Delta Q_n(n,x) = n\left[\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x\right] Q_{n-1}(n, x+1),$$

this relation being obtained by replacing N_x by $(N_x + D_x - D_{x-n})$ and D_x by D_{x-n} .

Formula XIII which was just derived is the general form of the relation we found to hold for the Poisson exponential function polynomials, i. e.

$$\Delta Q_n(x) = -nQ_{n-1}(x+1).$$

We find further that these polynomials satisfy a special form of (XI), i. e.

$$Q_{n+1}(x) + (x+n+1-\lambda)Q_n(x) - \lambda\Delta Q_n(x) = 0$$

and for formula (XII) we get the special form

$$Q_{n+1}(x) + (x+2n+1-\lambda)Q_n(x) + n(x+n)Q_{n-1}(x) = 0$$

This recurrence relation is also similar to the one given for Laguerre polynomials.

5. Turning now to the problem of obtaining a second order difference relation for the polynomials $Q_n(x)$, we proceed to difference formula (XI), i. e.

$$Q_{n+1}(x) = (N_x + D_x \;\; - D_{x+n})Q_n(x) + (N_x + D_x)\Delta Q_n(x)$$

and get

$$\Delta Q_{n+1}(x) = (\Delta N_x + \Delta D_x - \Delta D_{x+n}) Q_n(x)$$

$$+ (N_{x+1} + D_{x+1} - D_{x+n+1}) \Delta Q_n(x)$$

$$+ (\Delta N_x + \Delta D_x) \Delta Q_n(x) + (N_{x+1} + D_{x+1}) \Delta^2 Q_n(x).$$

Substituting for $\Delta Q_{n+1}(x)$ the value

$$(n+1)\left[\Delta N_x - \frac{n}{2}\Delta^2 D_x\right]\left[Q_n(x) + \Delta Q_n(x)\right]$$

found in formula (XIII), gives us

$$(n+1)\left[\Delta N_x - \frac{n}{2}\Delta^2 D_x\right]\left[Q_n(x) + \Delta Q_n(x)\right] =$$

$$\left[\Delta N_x + \Delta D_x - \Delta D_{x+n}\right] Q_n(x) + \left[N_{x+1} + D_{x+1} - D_{x+n+1}\right] \Delta Q_n(x)$$

$$+ \left[\Delta N_x + \Delta D_x\right] \Delta Q_n(x) + \left[N_{x+1} + D_{x+1}\right] \Delta^2 Q_n(x).$$

Collecting the coefficients of like terms and simplifying them, we finally get

$$\text{(XIV)} \qquad \begin{aligned} &(N_{x+1} + D_{x+1})\Delta^2 Q_n(x) + \left[N_{x-n+1} - (n-1)\Delta D_x\right] \Delta Q_n(x) \\ &- n\left[\Delta N_x - \frac{(n-1)}{2}\Delta^2 D_x\right] Q_n(x) = 0, \end{aligned}$$

a relation very similar in form to formula (IV) and consisting of the first and second differences of the polynomial $Q_n(x)$. This relation when applied to the Poisson exponential function gives

$$\lambda \Delta^2 Q_n(x) + (\lambda - x - 1)\Delta Q_n(x) + n Q_n(x) = 0,$$

an equation which can be checked by substituting the value of the general Poisson polynomial in it.

The extension of formula (XIV) to the polynomials $Q_n(m, x)$ and $Q_n(n, x)$ by making the proper substitutions for N_x

and D_x results in the following expressions:

$$
\begin{aligned}
&(N_{x+1}+D_{x+1})\Delta^2 Q_n(m,x)\\
(\mathrm{XIV}_m)\quad &+\left[N_{x-n+1}+D_{x-n+1}-D_{x-m-n+1}-(n-1)\Delta D_{x-m}\right]\Delta Q_n\,(m,x)\\
&-n\left[\Delta N_x+\Delta D_x-\Delta D_{x-m}-\left(\tfrac{n-1}{2}\right)\Delta^2 D_{x-m}\right]Q_n\,(m,x)=0,
\end{aligned}
$$

which may also be written as:

$$
\begin{aligned}
&(N_{x+1}+D_{x+1})\Delta^2 Q_n\,(m,x)\\
&+\left[N_{x-n+1}+(m-n+1)\Delta D_x-\frac{m(m+1)}{2}\Delta^2 D_x\right]\Delta Q_n\,(m,x)\\
&-n\left[\Delta N_x-\frac{n-2m-1}{2}\Delta^2 D_x\right]Q_n\,(m,x)=0
\end{aligned}
$$

In particular if $m=n$ we have:

$$
\begin{aligned}
&(N_{x+1}+D_{x+1})\Delta^2 Q_n\,(n,x)+\\
(\mathrm{XIV}_n)\quad &+\left[N_{x-n+1}+\Delta D_x-\frac{n(n+1)}{2}\Delta^2 D_x\right]\Delta Q_n\,(n,x)\\
&-n\left[\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x\right]Q_n\,(n,x)=0.
\end{aligned}
$$

6. The next set of relations we shall derive are recurrence relations for the polynomials $Q_n\,(m,x)$ and $Q_n\,(n,x)$ In the lemma proved in this chapter we found that

$$\Delta\left(D_{x-1}^{(m+1)}u_x\right)=\left[D_{x-1}^{(m)}u_x\right]\left[N_x+D_x-D_{x-m-1}\right].$$

Taking the n th difference of both sides of the equation gives:

$$
\begin{aligned}
\Delta^{n+1}\left(D_{x-1}^{(m+1)}u_x\right)=&\left(N_x+D_x-D_{x-m-1}\right)\Delta^n\left[D_{x-1}^{(m)}u_x\right]\\
&+n\left(\Delta N_x-\Delta D_x-\Delta D_{x-m-1}\right)\Delta^{n-1}\left[D_x^{(m)}u_{x+1}\right],
\end{aligned}
$$

the second difference of the trinomial $(N_x+D_x-D_{x-m-1})$ being equal to zero. Multiplying this last expression through by

$D^{(n+1)}_{x-m+n-1}$ and substituting for $D^{(n+1)}_{x-m+n-1}\Delta^{n+1}D^{(m+1)}_{x-1}u_x$ the value $D^{(m+1)}_{x-1}Q_{n+1}(m,x)u_x$, we get

$$D^{(m+1)}_{x-1}Q_{n+1}(m+1,x)u_x=(N_x+D_x-D_{x-m-1})D^{(m+1)}_{x-1}Q_n(m,x)u_x$$
$$+n(\Delta N_x+\Delta D_x-\Delta D_{x-m-1})D^{(m+2)}_x\left(\frac{N_x+D_x}{D_x}\right)Q_{n-1}(m,x+1)u_x.$$

Dividing through by $D^{(m+1)}_{x-1}u_x$ we get a recurrence relation involving the polynomials $Q_{n+1}(m+1,x)$, $Q_n(m,x)$ and $Q_{n-1}(m,x+1)$, i. e.

$$(\mathrm{XV}_m)\qquad \begin{aligned} Q_{n+1}(m+1,x)&=(N_x+D_x-D_{x-m-1})Q_n(m,x)\\ &+n\left[\Delta N_x+(m+1)\Delta^2D_x\right](N_x+D_x)Q_{n-1}(m,x+1). \end{aligned}$$

For $m=n$, this expression reduces to:

$$(\mathrm{XV}_n)\qquad \begin{aligned} Q_{n+1}(n+1,x)&=(N_x+D_x-D_{x-n-1})Q_n(n,x)\\ &+n\left[\Delta N_x+(n+1)\Delta^2D_x\right](N_x+D_x)Q_{n-1}(n,x+1). \end{aligned}$$

7. Another form of this relation is obtained by substituting the value found in (XIII_n) for $Q_{n-1}(n,x+1)$, i. e.

$$Q_{n-1}(n,x+1)=\frac{1}{n\left[\Delta N_x+\frac{n+1}{2}\Delta^2D_x\right]}\,\Delta Q_n(n,x),$$

in formula (XV_n), which gives

$$(\mathrm{XVI})\qquad \begin{aligned} Q_{n+1}(n+1,x)&=(N_x+D_x-D_{x-n-1})Q_n(n,x)\\ &+(N_x+D_x)\frac{\Delta N_x+(n+1)\Delta^2D_x}{\Delta N_x+\frac{(n+1)}{2}\Delta^2D_x}\,\Delta Q_n(n,x), \end{aligned}$$

a relation very similar to formula (VI).

8. There remains one more formula in Chapter II for which we have not yet found a parallel in this chapter, i. e. formula VII. To obtain this parallel expression, we difference formula (XVI), thereby obtaining:

$$\Delta Q_{n+1}(n+1,x) = (\Delta N_x + \Delta D_x - \Delta D_{x-n-1})\, Q_n(n,x)$$

$$+ (N_{x+1} + D_{x+1} - D_{x-n})\, \Delta Q_n(n,x)$$

$$+ \frac{\Delta N_x + (n+1)\Delta^2 D_x}{\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x} \left[(\Delta N_x + \Delta D_x)\Delta Q_n(n,x) + (N_{x+1} + D_{x+1})\Delta^2 Q_n(n,x) \right].$$

In formula (XIV_n) we found a value for

$$(N_{x+1} + D_{x+1})\Delta^2 Q_n(n,x)$$

which when substituted in this last expression gives us:

$$\Delta Q_{n+1}(n+1,x) = (\Delta N_x + \Delta D_x - \Delta D_{x-n+1})Q_n(n,x) + (N_{x+1} + D_{x+1} - D_{x-n})\Delta Q_n(n,x)$$

$$+ \frac{\Delta N_x + (n+1)\Delta^2 D_x}{\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x} \left[\Delta N_x + \Delta D_x + \frac{n(n+1)}{2}\Delta^2 D_x - N_{x-n+1} - \Delta D_x \right] \Delta Q_n(n,x)$$

$$+ n \frac{\Delta N_x + (n+1)\Delta^2 D_x}{\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x} \left[\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x \right] Q_n(n,x).$$

Collecting coefficients we get

$$\Delta Q_{n+1}(n+1,x) = \left[\Delta N_x + n\Delta N_x + \Delta D_x - \Delta D_{x-n-1} + n(n+1)\Delta^2 D_x \right] Q_n(n,x)$$

$$+ \left[N_{x+1} + D_{x+1} - D_{x-n} \right] Q_n(n,x)$$

$$+ \frac{\Delta N_x + (n+1)\Delta^2 D_x}{\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x} \left[\Delta N_x - N_{x-n+1} + \frac{n(n+1)}{2}\Delta^2 D_x \right] \Delta Q_n(n,x)$$

and by simplifying the coefficients this expression finally reduces to the formula

$$
\text{(XVII)}\quad
\begin{aligned}
\Delta Q_{n+1}(n+1,x) &= (n+1)\left[\Delta N_x+(n+1)\Delta^2 D_x\right] Q_n(n,x) \\
&+\Bigg\{N_{x+1}+(n+1)\Delta D_x-\frac{n(n+1)}{2}\Delta^2 D_x \\
&-\left[\frac{\Delta N_x+(n+1)\Delta^2 D_x}{\Delta N_x+\left(\frac{n+1}{2}\right)\Delta^2 D_x}\right]\left[N_{x-n}-\frac{n(n+1)}{2}\Delta^2 D_x\right]\Bigg\}\Delta Q_n(n,x)
\end{aligned}
$$

a relation which is also similar in form to formula VII.

Before concluding this chapter, we might examine the character of the polynomials $Q_n(n,x)$ when the original function is the Poisson exponential function $\psi(x)=\frac{e^{-\lambda}\lambda^x}{x!}$.

We find these polynomials to have the following form:

$$\frac{1}{\psi(x)}\Delta\frac{xe^{-\lambda}\lambda^x}{x!}=Q_1(1,x)=\lambda-x,$$

$$\frac{1}{\psi(x)}\Delta^2\frac{x^{(2)}e^{-\lambda}\lambda^x}{x!}=Q_2(2,x)=\lambda^2-2\lambda x+x^{(2)}$$

$$\frac{1}{\psi(x)}\Delta^3\frac{x^{(3)}e^{-\lambda}\lambda^x}{x!}=Q_3(3,x)=\lambda^3-3\lambda^2x+3\lambda x^{(2)}-x^{(3)},$$

- -

$$\frac{1}{\psi(x)}\Delta^n\frac{x^{(n)}e^{-\lambda}\lambda^x}{x!}=Q_n(n,x)=\lambda^n-n\lambda^{n-1}x+\frac{n(n-1)}{2!}\lambda^{n-2}x^{(2)}+\cdots+(-1)^n x^{(n)},$$

$$=\left[\lambda-x^{(1)}\right]^n=Q_n(x-n).$$

Substituting the proper values for N_x and D_x in formula (XIV_n) we get

$$\lambda\Delta^2Q_n(n,x)+(\lambda-x+n-1)\Delta Q_n(n,x)+n\,Q_n(n,x)=0$$

In the same way we find for formula (XI_n) the relation

$$Q_{n+1}(n+1,x) = (\lambda - x + n)Q_n(n,x) + \lambda \Delta Q_n(n,x)$$

and for formula (XVII), the reduced relation

$$\Delta Q_{n+1}(n+1,x) = -(n+1)Q_n(n,x),$$

which is somewhat like the relation obtained for ($XIII_n$).

We might call attention to the fact that these polynomials are identical with the polynomials obtained by Charlier[1] satisfying the relations

$$\delta^{-1}\left[T_n(x)\,\delta^m \frac{e^{-\lambda}\lambda^x}{x!}\right]_{-\infty}^{+\infty} \begin{array}{l} = 0 \text{ for } m \neq n \\ = 1 \text{ for } m = n \end{array}$$

9. Summarizing the results of this chapter, we have found that if the general solution $g(x)$ of the difference equation

$$\Delta u_x = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2}\, u_x$$

is used as the generating function $g(x)$ in the Charlier Type B series, that the successive differences give rise to two general types of polynomials which we defined as follows:

$$Q_n(x) = \frac{1}{u_x} D_x^{(n)} \Delta^{n+1} u_x$$

and

$$Q_n(n,x) = \frac{1}{u_x} \Delta^n D_{x-n}^{(n)}\, u_x\,.$$

With the aid of the properties of the Δ operator, we derived a set of relations and equations for these polynomials of the following form:

[1]C. V. L. Charlier: "Ueber die Darstellung willkürlicher Funktionen," p. 34.

(XI) $$Q_{n+1}(x)=(N_x+D_x-D_{x+n})Q_n(x)+(N_x+D_x)\Delta Q_n(x),$$

(XI$_n$) $$Q_{n+1}(n+1,x)=(N_x+\Delta D_{x-1})Q_n(n+1,x)+(N_x+D_x)\Delta Q_n(n,x),$$

(XII) $$\begin{aligned}Q_{n+1}(x)&=(N_{x+n}-n\Delta D_{x+n-1})Q_n(x)\\&\quad+nD_{x+n-1}\left[\Delta N_x-\frac{(n-1)}{2}\Delta^2 D_x\right]Q_{n-1}(x),\end{aligned}$$

(XII$'$) $$\begin{aligned}Q_{n+1}(x)&=(N_x+D_x-D_{x+n})Q_n(x)\\&\quad+n\left[\Delta N_x-\frac{(n-1)}{2}\Delta^2 D_x\right](N_x+D_x)Q_{n-1}(x+1),\end{aligned}$$

(XIII) $$\Delta Q_n(x)=n\left[\Delta N_x-\frac{(n-1)}{2}\Delta^2 D_x\right]Q_{n-1}(x+1),$$

(XIII$_n$) $$\Delta Q_n(n,x)=n\left[\Delta N_x+\frac{n+1}{2}\Delta^2 D_x\right]Q_{n-1}(n,x+1),$$

(XIV) $$\begin{aligned}&(N_{x+1}+D_{x+1})\Delta^2 Q_n(x)+\left[N_{x-n+1}-(n-1)\Delta D_x\right]\Delta Q_n(x)\\&\quad-n\left[\Delta N_x-\frac{(n-1)}{2}\Delta^2 D_x\right]Q_n(x)=0,\end{aligned}$$

(XIV$_n$) $$\begin{aligned}&(N_{x+1}+D_{x+1})\Delta^2 Q_n(n,x)+\left[N_{x-n+1}+\Delta D_x-\frac{n(n+1)}{2}\Delta^2 D_x\right]\Delta Q_n(n,x)\\&\quad-n\left[\Delta N_x+\frac{n+1}{2}\Delta^2 D_x\right]Q_n(n,x)=0,\end{aligned}$$

(XV$_n$) $$\begin{aligned}Q_{n+1}(n+1,x)&=(N_x+D_x-D_{x-n-1})Q_n(n,x)\\&\quad+n\left[\Delta N_x+(n+1)\Delta^2 D_x\right](N_x+D_x)Q_{n-1}(n,x+1),\end{aligned}$$

(XVI) $$\begin{aligned}Q_{n+1}(n+1,x)&=(N_x+D_x-D_{x-n-1})Q_n(n,x)\\&\quad+(N_x+D_x)\frac{\Delta N_x+(n+1)\Delta^2 D_x}{\Delta N_x+\frac{(n+1)}{2}\Delta^2 D_x}\Delta Q_n(n,x),\end{aligned}$$

$$\text{(XVII)}\quad \Delta Q_{n+1}(n+1,x) = (n+1)\left[\Delta N_x + (n+1)\Delta^2 D_x\right] Q_n(n,x)$$
$$+\Bigg\{ N_{x+1} + (n+1)\Delta D_x - \frac{n(n+1)}{2}\Delta^2 D_x$$
$$-\left[\frac{\Delta N_x + (n+1)\Delta^2 D_x}{\Delta N_x + \frac{(n+1)}{2}\Delta^2 D_x}\right]\left[N_{x-n} - \frac{n(n+1)}{2}\Delta^2 D_x\right]\Bigg\}\Delta Q_n(n,x)$$

Each of these formulas corresponds and is similar to a formula found in Chapter II. In fact, it seems probable that if we developed the formulas in this present chapter from the equation

$$\frac{\Delta u_x}{\Delta x} = \frac{N_x}{D_x} u_x$$

and permitted the Δ_x to approach zero as a limit, the formulas of Chapter II would result, the above formulas being the case where $\Delta_x = 1$.

E. H. Hildebrandt

A NEW FORMULA FOR PREDICTING THE SHRINKAGE OF THE COEFFICIENT OF MULTIPLE CORRELATION

By

DR. R. J. WHERRY
Cumberland University, Lebanon, Tennessee

With the perfection of the Doolittle Method for the solution of the constant values necessary for the multiple correlation and prediction technique, we may expect a constant increase in the use of this method in statistical practice. Theoretical statisticians have recognized for some time however that the multiple correlation coefficient, derived from a large number of independent variables, is apt to be deceptively large due to chance factors. When prediction equations derived in this manner are applied to subsequent sets of data, there is apt to be a rather large shrinkage in the resulting correlation coefficient obtained, as compared with the original observed multiple correlation coefficient. In order to avoid over optimism it is necessary to have some equation which will predict the most probable value of this shrinkage. The development of such a formula is the purpose of this paper.

The most promising formula of this type so far developed is the B. B. Smith formula, presented by M. J. B. Ezekial at the December, 1928, meeting of the American Mathematical Society held at Chicago. This formula is

$$(1) \qquad \bar{R}^2 = 1 - \frac{1-R^2}{1-\frac{M}{N}} = \frac{NR^2 - M}{N-M}$$

where $\bar{R}$ = the estimated correlation obtaining in the universe
R = the observed multiple correlation coefficient
M = the number of *independent* variables
N = the number of observations (the statistical population).

This formula was evidently developed by B. B. Smith by an application of the method of least squares as follows (the derivation is that of the author, since he could not find it given elsewhere):

The customary formula for the coefficient of multiple correlation may be written in the form

(2) $$R^2 = 1 - \frac{S_0^2}{\sigma_0^2}$$

where

(3) $$S_0^2 = \frac{\Sigma v^2}{N}$$

where

(4) $$v = X_0 - \bar{X}_0$$

The method of least squares, however, says that the most probable value of the standard error of estimate is not that given in equation (3) but

(5)[1] $$\bar{S}_0^2 = \frac{\Sigma v^2}{N-M} = \frac{N}{N-M} \cdot S_0^2$$

Now, if we substitute the value of (5) in place of (3) in equation (2), we have at once

[1]See Merriman, Method of Least Squares, John Wiley & Sons, London, 8th Edition, pp. 80-82. Also see derivation later in this paper.

$$(6) \qquad \bar{R}^2 = 1 - \frac{s_o^2}{\sigma_o^2} \cdot \frac{N}{N-M}$$

and since, by (2) above, we have $\frac{s_o^2}{\sigma_o^2}$ equal to $(1 - R^2)$, we have

$$(7) \qquad \bar{R}^2 = 1 - \frac{N(1-R^2)}{N-M} = 1 - \frac{1-R^2}{1-\frac{M}{N}}$$

which is, exactly, the B. B. Smith formula (1).

This formula has been widely used during the last few years, but up until recently had not been subjected to much critical examination. However, in a recent article in the *Journal of Educational Psychology*[1], S. C. Larson actually tested the formula empirically on some data obtained from the Mississippi Survey conducted by M. V. O'Shea, obtaining the results indicated in the tables and graphs below, and on the basis of which he reached the following conclusion:

"The Smith Shrinkage-Reduction formula parallels all of the empirical findings but quite consistently gives values which are in excess of those obtained under present experimental conditions." This meant that the Smith formula predicted shrinkages consistently greater than those actually obtained.

It was in view of this reported empirical difference that the writer started his attempt to derive the Smith formula and hit on the method given above. The question at once arose in the writer's mind as to why, when the standard error of estimate had been corrected to correspond to the most probable value by a least squares criterion, the standard deviation of the dependent variable had not been treated in the same fashion.

---

[1]"The Shrinkage of the Coefficient of Multiple Correlation," Jan., 1931, pp. 45-55.

Merriman, whose formula we used above in correcting the standard error of estimate (5), likewise, and by identical reasoning, shows that the most probable value of the standard deviation of the dependent variable existing in the universe, should really be represented by the following relationships:

Where

$$(8) \qquad \sigma_o^2 = \frac{\Sigma x_o^2}{N}$$

we find

$$(9) \qquad \bar{\sigma}_o^2 = \frac{\Sigma x_o^2}{N-1} = \sigma_o^2 \cdot \frac{N}{N-1}$$

which reduces formula (6) to the form

$$(10a) \qquad \bar{R}^2 = 1 - \frac{s_o^2}{\sigma_o^2} \frac{\frac{N}{N-M}}{\frac{N}{N-1}}$$

and when the same substitution is made as in step (7) above, we have

$$(10b) \qquad \bar{R}^2 = \frac{(N-1)R^2-(M-1)}{N-M}$$

which is, by a more correctly applied criterion of least squares, the formula we have been seeking, and is a closer approximation than that given by the Smith formula.

The reasons for the substitutions made above in our formulae may not be entirely clear to all readers, so we now present the derivations of the formulae given in (5) and (9) above. The derivations given here are directly adapted from those of Merriman referred to above, but have been translated into the customary statistical notation whenever possible.

First, let us consider the derivation of the value in (9). As

stated in (8) the most customary form of Sigma is

$$\sigma_o^2 = \frac{\Sigma x_o^2}{N} \tag{8}$$

where

$$x_o = x - M_x . \tag{11}$$

Each value x_o has a certain error, however, due to the fact that the value of the mean is merely the most probable value, not the true value. So for each x_o value there is a small unknown error δx_o, so that if we take $\bar{x}_o$ to be the true value of a deviation we have

$$\bar{x}_o = x_o + \delta x_o \tag{12}$$

and, squaring and summating, disregarding the terms involving second power delta terms as small in comparison with the first power terms, we have

$$\Sigma \bar{x}_o^2 = \Sigma x_o^2 + 2\Sigma x_o \delta x_o \tag{13}$$

Now, by the laws of probability, we know that the probability of the occurrence of an error $\bar{x}_o$, whose measure of precision is "h," is

$$\Pi = h d\bar{x} \pi^{-\frac{1}{2}} e^{-\bar{x}^2 h^2} \tag{14}$$

multiplying both sides of this equation by $\bar{x}^2$ and summating between the limits plus and minus infinity, we have

$$\Sigma \Pi \bar{x}^2 = \int_{-\infty}^{+\infty} h \bar{x}^2 \pi^{-\frac{1}{2}} e^{-h^2 \bar{x}^2} d\bar{x} = \frac{1}{2h^2} \tag{15}$$

and since $\Sigma \Pi \bar{x}^2$ is the same as $\frac{\Sigma \bar{x}^2}{N}$, since in our work we assume the weight of each value $\bar{x}$, for each of the N observations, to be $\frac{1}{N}$, we have

(16) $$\frac{\Sigma \bar{x}^2}{N} = \frac{1}{2h^2}$$

or

(16a) $$\Sigma \bar{x}^2 = \frac{N}{2h^2}$$

Likewise, if we let

(17) $$2\Sigma x_o \delta x_o = u^2$$

the probability of the system of errors, u^2, is

(18) $$\Pi' = h\, du\, \pi^{-\frac{1}{2}} e^{-u^2h^2}$$

and the mean of all of the possible values of u^2 is

(19) $$\frac{h}{\pi^{\frac{1}{2}}} \int_{-\infty}^{+\infty} u^2 e^{-h^2u^2} du = \frac{1}{2h^2}$$

and this must be taken as the best attainable value of u^2. But it was shown that the quantity $\frac{1}{2h^2}$ is equal to $\frac{\Sigma \bar{x}^2}{N}$ (16). Hence

(20) $$\Sigma \bar{x}^2 = \Sigma x^2 + \frac{\Sigma \bar{x}^2}{N}$$

from which

(9) $$\bar{\sigma}_o^2 = \frac{\Sigma \bar{x}^2}{N} = \frac{\Sigma x^2}{N-1} = \sigma_o^2 \cdot \frac{N}{N-1}$$

which was to be proved.

To derive (5) we proceed in much the same manner. After our normal equations have been solved for the most probable values of $\beta_{01}, \beta_{02}, \beta_{03}, \cdot \beta_{0m}$ for our set of data, we know that these are not the true values, but that they err by small unknown corrections $\delta\beta_{01}, \delta\beta_{02}, \delta\beta_{03}, \quad \delta\beta_{0m}$, the corresponding true values for the universe being $(\beta_{01} + \delta\beta_{01})$, $(\beta_{02} + \delta\beta_{02})$, $(\beta_{03} + \delta\beta_{03})$, . . . $(\beta_{0m} + \delta\beta_{0m})$.

Now, if we substitute the most probable values of the Betas in our original observation equations, they will not reduce to zero, bus will leave small residuals v_1, v_2, ... v_N, thus

$$\bar{x}_{0_1} - x_{0_1} = \beta_{01} x_{1_1} + \beta_{02} x_{2_1} + \beta_{03} x_{3_1} + \quad + \beta_{0m} x_{m_1} - x_{0_1} = v_1$$

$$\bar{x}_{0_2} - x_{0_2} = \beta_{01} x_{1_2} + \beta_{02} x_{2_2} + \beta_{03} x_{3_2} + \cdot \quad + \beta_{0m} x_{m_2} - x_{0_2} = v_2$$

$$\cdot \qquad \cdot \qquad \cdot \qquad \cdot$$

$$\bar{x}_{0_M} - x_{0_M} = \beta_{01} x_{1_N} + \beta_{02} x_{2_N} + \beta_{03} x_{3_N} + \cdot \cdot \quad + \beta_{0m} x_{m_N} - x_{0_N} = v_N$$

while if the corresponding true values be substituted, we obtain

$$(\beta_{01} + \delta\beta_{01}) x_{1_1} + (\beta_{02} + \delta\beta_{02}) x_{2_1} + \cdot \quad + (\beta_{0m} + \delta\beta_{0m}) x_{m_1} - x_{0_1} = \bar{v}_1$$

$$(\beta_{01} + \delta\beta_{01}) x_{1_2} + (\beta_{02} + \delta\beta_{02}) x_{2_2} + \quad + (\beta_{0m} + \delta\beta_{0m}) x_{m_2} - x_{0_2} = \bar{v}_2$$

$$\cdot \quad \cdot \quad \cdot \qquad \cdot \quad \cdot \quad \cdot$$

$$(\beta_{01} + \delta\beta_{01}) x_{1_N} + (\beta_{02} + \delta\beta_{02}) x_{2_N} + \cdot \quad + (\beta_{0m} + \delta\beta_{0m}) x_{m_N} - x_{0_N} = \bar{v}_N$$

Subtracting each of the former equations from the latter, we obtain

$$v_1 + \delta\beta_{01}\, x_{1_1} + \delta\beta_{02}\, x_{2_1} + \qquad + \delta\beta_{0m}\, x_{m_1} = \bar{v}_1$$

$$v_2 + \delta\beta_{01}\, x_{1_2} + \delta\beta_{02}\, x_{2_2} + \quad \cdot \qquad + \delta\beta_{0m}\, x_{m_2} = \bar{v}_2$$

$$\cdot \quad \cdot \qquad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$v_N + \delta\beta_{01}\, x_{1_N} + \delta\beta_{02}\, x_{2_N} + \quad \cdot \qquad + \delta\beta_{0m}\, x_{m_N} = \bar{v}_N$$

Now the principle of least squares provides that $\Sigma\, \bar{v}^2$ shall be made a minimum to give the most probable values of β_{01} , β_{02} , β_{0m} , and by the solution of the normal equations by the Doolittle method its minimum value is found to be $\Sigma\, v^2$ From the residual equations we may find the relationship existing between the values $\Sigma\, \bar{v}^2$ and $\Sigma\, v^2$. Thus, if we square each equation immediately above and then summate we have (if we neglect squares and products of the delta values as small in comparison with the first powers).

$$(21) \quad \Sigma v^2 + 2\,\delta\beta_{01}\,\Sigma x_1 v + 2\delta\beta_{02}\,\Sigma x_2 v + \qquad + 2\,\delta\beta_{0m}\,\Sigma x_m v = \Sigma \bar{v}^2$$

which we may write as

$$(22) \quad \Sigma v^2 + u_1^2 + u_2^2 + \cdot \qquad \cdot + u_m^2 = \Sigma \bar{v}^2$$

Now, by analogous reasoning to that in steps (14), (15), and

(16), we may set

$$\Sigma \bar{v}^2 = \frac{N}{2h^2} \tag{23}$$

Further, if there be but one independent variable, there will be but one $2\delta\beta_{ox}\Sigma x_x v$ and its value by the same process used in steps (18) and (19) can be shown to be

$$u_x^2 = \frac{1}{2h^2} \tag{24}$$

and since that is true whichever unknown quantity be considered, the values of each u_x^2 value must be $\frac{1}{2h^2}$; and as there are M of these values the above equation (22) becomes

$$\Sigma v^2 + \frac{M}{2h^2} = \frac{N}{2h^2}$$

from which

$$h = \sqrt{\frac{N-M}{2\Sigma v^2}} \tag{25}$$

Therefore, from the constant relationship which exists between the value "h" and the Probable Error, we have

$$PE_{\bar{v}} = 0\ 6745\sqrt{\frac{\Sigma v^2}{N-M}} \tag{26}$$

and therefore, by the relationship existing between the probable error and the standard deviation we have at once

$$\sigma_{\bar{v}}^2 = \bar{s}_o^2 = \frac{\Sigma v^2}{N-M} \tag{5}$$

which was to be proved

The next step was to test out the formula empirically. This was done by using Larson's material, with the results indicated in the tables below, and in the graphs which show the same

set of facts, but which make the results much more apparent.

An inspection of the tables and graphs will show at once that the new formula predicts what will actually happen much more accurately than the Smith formula did. In graph 1, for example, the agreement is so good that the results appear almost to have been a regression line fit to the particular set of data.

It was to have been expected that if the formula actually predicted the most probable values of the correlations obtaining in the universe that the errors incurred by the use of the formula would be normally distributed around zero as a mean value. Graph 3 presents a comparison of the error curves obtained by use of the Smith and the Wherry prediction formulae, together with an approximation to the normal curve. As a further and more scientific check the criteria for a normal curve as set forth by Rietz[1] were applied to the data. His criteria are

$$\mu_1 = 0, \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0, \ \beta_2 = \frac{\mu_4}{\mu_2^2} = 3, \quad \text{where } \mu_n = \frac{\Sigma x^n}{N}$$

The results for the two formulae are given below.

(Results based on an expectancy of zero)

	Smith Formula	Wherry Formula
μ_1	.00138	.00038
β_1	223	025
β_2	3.004	3.703

[1]Rietz, H. L Mathematical Statistics, Carus Mathematical Monograph No. 3, Mathematical Association of America, Chicago 1927, pp. 58-59.

It is apparent therefore that the Wherry formula gave much better results for both the first criterion (mean error) and the second criterion (skewness), but that the excess was greater for the Wherry formula than for the Smith formula. However, one cannot quarrel too much with getting errors actually smaller than would be expected by assuming normality. Even this superiority is seen to be fictitious if the distributions are measured from their own means rather than from an expected mean of zero. When this is done, which is the manner in which the criteria are customarily used, we have

(Results based on means of distributions)

	Smith Formula	Wherry Formula
α_1	.000	.000
β_1	1.712	.025
β_2	5.524	3.753

Thus, we find that the Smith distribution has, in reality, even a greater excess than does the Wherry formula, but has it at a point farther removed from the desired value.

SUMMARY AND CONCLUSIONS

1. Larson has shown that the theoretically expected shrinkage is an empirical fact

2. Larson has shown that the Smith formula, when tested empirically, consistently over-estimates this shrinkage as determined empirically.

3. It has been demonstrated that the new Wherry formula,

both by a least squares criterion and by actual application, is more nearly true than the corresponding Smith formula.

4 The correct formula for the shrunken coefficient of multiple correlation is

$$\bar{R}^2 = \frac{(N-1)\,R^2 - (M-1)}{N-M}$$

where $\bar{R}$ = the estimated correlation obtaining in the universe

R = the observed coefficient of multiple correlation

M = the number of *independent* variables

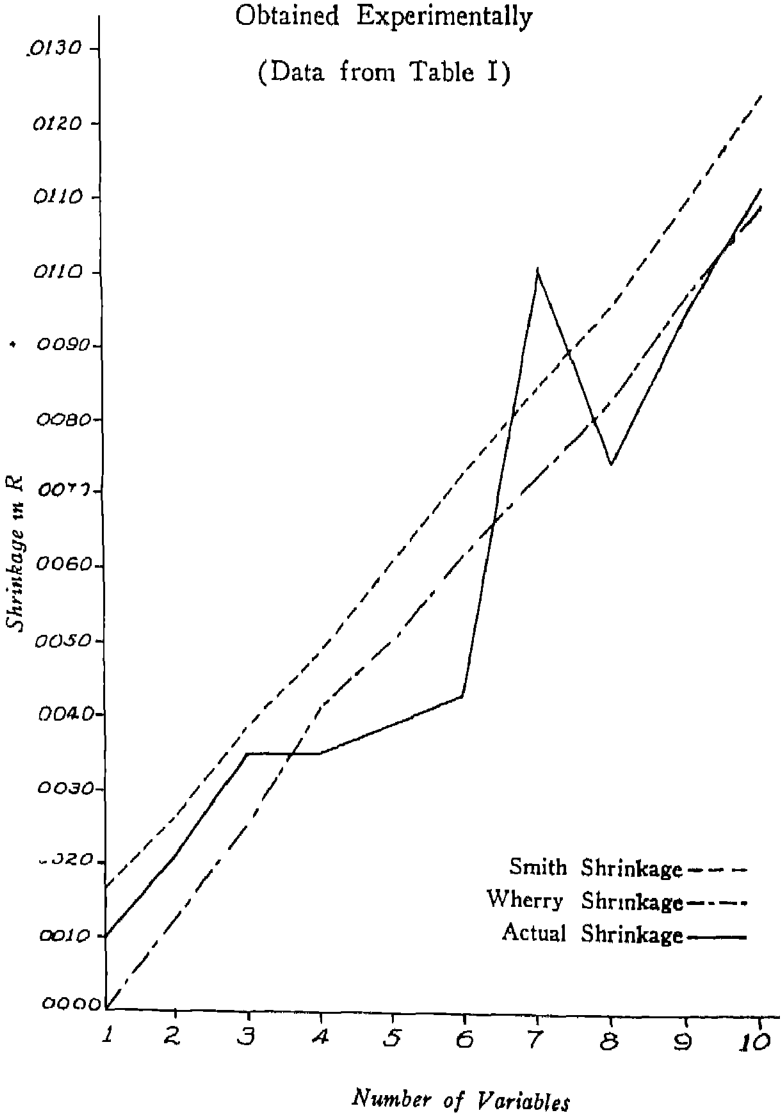
and N = the number of observations (statistical population)

R. J. Wherry

GRAPH 1.

Shrinkage as Obtained by Use of the Formulae and Also as Obtained Experimentally

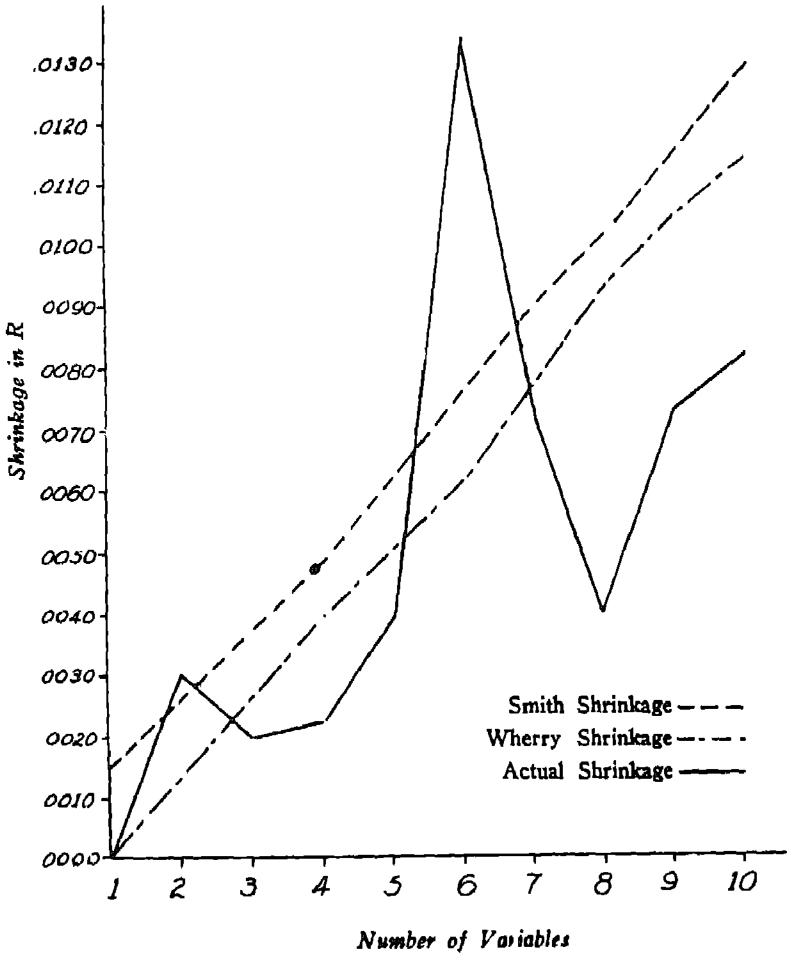
(Data from Table I)



GRAPH 2

Shrinkage as Obtained by Use of the Formulae and Also as Obtained Experimentally

(Data from Table II)



GRAPH 3

Ogive Showing the Distribution of Error in Predicting Shrinkage

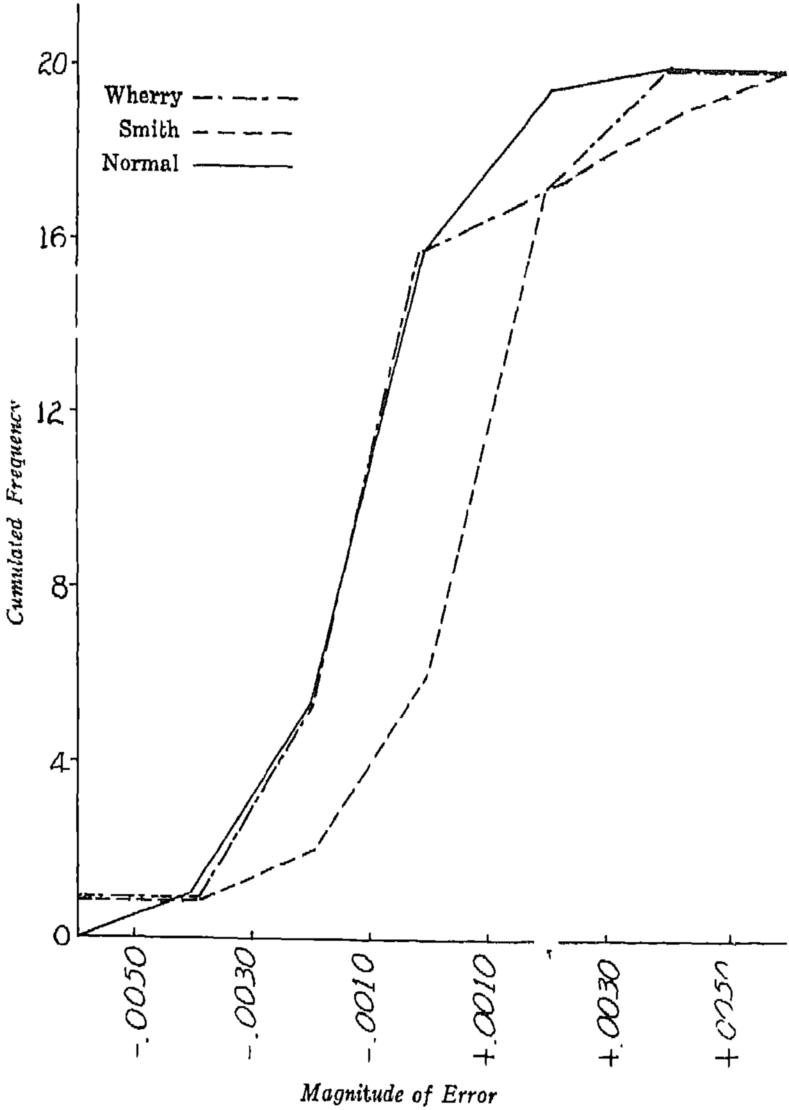


TABLE I*

Showing the Actual Shrinkage in R Found When the Prediction Equation Found on One Group of Subjects Is Applied to a Comparable Group, Together with the Shrinkage of R as Indicated by the Smith and Wherry Formulae. The Statistical Population (n) Is 200 Throughout.

m	1	2	3	4	5	6	7	8	9	10
R	7042	,7794	.7834	7872	7880	7907	7929	.7941	.7944	7945
Actual Shrinkage	0000	.0021	.0036	0036	.0060	0044	.0102	.0075	.0097	.0113
Shrinkage by Smith formula	0017	.0026	.0038	0049	0062	0074	.0085	0097	.0110	.0123
Shrinkage by Wherry formula	.0000	.0013	0025	.0040	0049	.0062	0073	.0085	.0098	0111

*The article by Larson reported the values for the Smith formula erroneously, due to a misconception of the meaning of m. Those in the present tables are the correct values.

TABLE II

Showing for a Second Set of Groups the Same Facts as Obtain in Table I

m	1	2	3	4	5	6	7	8	9	10
R	.7402	.7759	.7813	.7826	.7847	.7858	.7859	.7863	.7868	.7869
Actual Shrinkage	.0000	.0031	.0019	.0023	.0041	.0133	.0073	.0042	.0074	.0083
Shrinkage by Smith formula	.0015	.0026	.0038	.0051	.0063	.0076	.0089	.0102	.0115	.0129
Shrinkage by Wherry formula	.0000	.0013	.0026	.0038	.0051	.0063	.0076	.0089	.0105	.0115

TABLE III

Showing the Mean Error Attained by the Use of the Smith and Wherry Shrinkage Formulae.

Formula	Table I	Table II	Tables I and II
Smith	00097	00180	00138
Wherry	.00018	00057	00038
N	10	10	20

THE USE OF THE RELATIVE RESIDUAL IN THE APPLICATION OF THE METHOD OF LEAST SQUARES

By

WALTER A. HENDRICKS
Junior Biologist, Bureau of Animal Industry,
U. S. Department of Agriculture.

The method of least squares offers a precise method of fitting a curve describing the relation between two or more related, measurable variables, but certain criteria must be fulfilled to justify its application. First, the type of equation selected for fitting must be the true mathematical expression of the law governing the relationship of the variables. Secondly, all errors of measurement, made in obtaining the observed values of the variables when the data were collected, must be distributed according to the well-known laws of probability.[1]

This paper is concerned with the latter of these two criteria. The fundamental theory upon which the method of least squares is based can be found in any text-book on the subject and need not be elaborated upon here. However, it may be well to point out a very pertinent, if somewhat elementary, aspect of the theory which facilitates the ready visualization of the fundamental concepts involved.

[1]Steinmetz, C. P. Engineering Mathematics. McGraw-Hill Book Co., New York (1917).

The application of the method of least squares to curve fitting, as ordinarily described in works on the subject, is perfectly analogous to the calculation of the arithmetic mean of a number of measurements made upon a single, constant quantity This may be easily demonstrated as follows:

Let $Y = f(X)$ describe the relation existing between an independent variable, X, and a dependent variable, Y If it is desired to find the most probable value of the dependent variable when X has some definite value, X_1, the most direct method of procedure would be to make a number of measurements of Y at this value of X and calculate their arithmetic mean, provided, of course, that the errors of measurement were distributed according to the laws of probability in a normal frequency distribution According to the elementary theory of statistics, the most probable value of the dependent variable, $f(X_1)$, would be such that the sum of the squares of the deviations of the actual measurements from this value would be a minimum.

If X is conceived to be varying in value so rapidly that it is impossible to make more than one measurement of Y at any value of X, this direct method can not be employed. However, the most probable value of $f(X_1)$ can still be determined. Let $Y_1, Y_2, Y_3, \ldots Y_n$ each represent a measured value of Y at values $X_1, X_2, X_3, \ldots X_n$, respectively, of the independent variable Since the errors of measurement are assumed to be distributed according to the law of chance, an error of a given magnitude is equally likely to occur at any value of X. In other words, exactly the same errors would be made in obtaining one measurement of each of the quantities $Y_1, Y_2, Y_3, \ldots Y_n$ as if $f(X_1)$ were measured n times. These errors may, therefore, be considered as having been made in measuring a single, constant quantity. Therefore, if $f(X)$ denotes the most probable value of Y at any value of X and Y denotes the corresponding observed value, the most probable values of the dependent variable which can be calculated from any set of

data are such that the sum of the squares of the differences, $f(X)-Y$, is a minimum.

It is important to bear in mind that this conception of the distribution of errors of measurement is justified only when an error of a given magnitude is equally likely to occur at any value of X. In actual practice it often happens that this ideal condition is not realized The magnitude of the errors of measurement is often influenced by the magnitude of the quantity which is being measured. In obtaining the live weights of animals at different ages, for example, it is common practice to use a less delicate balance in making the weighings as the animals become larger, and the magnitude of the errors of measurement increases as the sensitivity of the balance decreases. Other factors which tend to increase the magnitude of the errors may also be in operation. The error, or rather the unreliability, of the weight of a 1,000 pound steer would be greater than that of a 100 pound calf, even though an equally sensitive balance were used in making both weighings, because of a greater content of material in the digestive tract and excretory organs and the increased effect of the movements of the animal.

It is highly probable that in many fields of investigation such disturbing influences are encountered more frequently than the ideal conditions which justify the application of the method of least squares as ordinarily described.

Pearl and Reed recognized the need for modifying the application of the method of least squares to compensate for changes in the probability of the occurrence of an error of any given magnitude and suggested, as stated by Pearl,[1] that it would be more logical in many instances to employ residuals of the type $\frac{f(X)-Y}{Y}$. The use of such residuals was based on the assumption that if the errors of measurement were expressed as percentages of the

[1]Pearl, Raymond. Studies in Human Biology. Williams & Wilkins, Baltimore (1924)

magnitude of the quantities measured, the percentage errors would be distributed at random according to the law of probability. In many practical problems this assumption appears to be justifiable.

The study herein reported was made to determine the extent of the error made when the method of least squares as ordinarily described is applied to data in which the percentage, rather than the absolute errors of measurement are distributed according to the law of chance.

The writer desired a hypothetical set of errors of measurement which, when expressed as percentages of the quantities measured, would come as near as possible to forming a normal frequency distribution.

TABLE I

Ideal Frequency Distribution of 41 Throws of 12 Dice in Which a Throw of 4, 5, or 6 Points Is Considered a Success.

SUCCESSES	FREQUENCY
2	1
3	2
4	5
5	8
6	9
7	8
8	5
9	2
10	1
	—
Total	41

Mills[1] gives the results of fitting a normal frequency curve to Weldon's distribution of 4096 throws of 12 dice, described by Yule,[2] in which a throw of 4, 5, or 6 points was considered a success. If each frequency, calculated from the fitted curve, is divided by 100 and the results rounded off to whole numbers, the frequency distribution given in Table I is obtained.

If hypothetical errors of measurement are substituted for

TABLE II

Ideal Frequency Distribution of 41 Hypothetical Percentage Errors of Measurement.

ERROR (Per cent of quantity measured)	FREQUENCY
+ 8	1
+ 6	2
+ 4	5
+ 2	8
0	9
− 2	8
− 4	5
− 6	2
− 8	1
Total	41

[1]Mills, F. C Statistical Methods Applied to Economics and Business. Henry Holt & Co., New York (1924).

[2]Yule, G. Udny. Introduction to the Theory of Statistics. Charles Griffin & Co., Ltd., London (1927).

successes in this frequency table, the resulting distribution may be considered to represent a distribution of random errors of measurement which might be made in obtaining a series of 41 measurements of a variable. The most probable error should obviously be zero. If the total range in magnitude of the errors is assumed to be from + 8 per cent to −8 per cent and the precision of measurement is such that each error differs from the next larger or smaller error by 2 per cent, the distribution of these hypothetical errors of measurement should be as given in Table II

From the simple equation, $Y = 100\,X^2$, 41 values of Y were calculated, using values of X from 1 to 41, inclusive. Each calculated value of Y was then changed by algebraically subtracting the hypothetical errors of measurement given in Table II. All the percentage errors of each magnitude were arbitrarily distributed as uniformly as possible throughout the data. These altered values of Y will hereafter be termed the "observed" values and the original values, from which they were calculated, the "true" values. The observed values of Y, together with the true values and the assumed errors of measurement from which they were calculated, are given in Table III.

In order to be certain that the errors were actually distributed in such a manner that the probability of the occurrence of a percentage error of any given magnitude was the same at all values of X, the writer employed Pearson's method of square contingency as described by Yule.[1] A 16-cell contingency table was constructed in which the percentage errors were classified according to the values of X at which they occurred The chi-square test for contingency was applied to this table.

Table IV shows the actual distribution of the percentage errors, together with the corresponding theoretical frequencies. Since there are 4 rows and 4 columns of cells in the table, the number of algebraically independent differences between theoret-

[1]Loc. cit.

TABLE III

Calculation of the Observed Values of Y from the True Values.

X	$100X^2$	ERROR		Y	X	$100X^2$	ERROR		Y
		Per cent	Actual Units				Per cent	Actual Units	
1	100	− 2	− 2	102	22	48400	+ 8	+ 3872	44528
2	400	+ 4	+ 16	384	23	52900	0	0	52900
3	900	0	0	900	24	57600	− 4	− 2304	59904
4	1600	− 4	− 64	1664	25	62500	− 2	− 1250	63750
5	2500	+ 6	+ 150	2350	26	67600	0	0	67600
6	3600	− 6	− 216	3816	27	72900	− 2	− 1458	74358
7	4900	+ 2	+ 98	4802	28	78400	+ 6	+ 4704	73696
8	6400	0	0	6400	29	84100	+ 2	+ 1682	82418
9	8100	− 2	− 162	8262	30	90000	+ 4	+ 3600	86400
10	10000	+ 2	+ 200	9800	31	96100	− 4	− 3844	99944
11	12100	+ 4	+ 484	11616	32	102400	+ 2	+ 2048	100352
12	14400	− 4	− 576	14976	33	108900	0	0	108900
13	16900	− 8	− 1352	18252	34	115600	+ 2	+ 2312	113288
14	19600	+ 2	+ 392	19208	35	122500	− 2	− 2450	124950
15	22500	− 2	− 450	22950	36	129600	0	0	129600
16	25600	0	0	25600	37	136900	+ 4	+ 5476	131424
17	28900	+ 2	+ 578	28322	38	144400	− 6	− 8664	153064
18	32400	+ 4	+ 1296	31104	39	152100	− 2	− 3042	155142
19	36100	− 2	− 722	36822	40	160000	0	0	160000
20	40000	0	0	40000					
21	44100	+ 2	+ 882	42318	41	168100	− 4	− 6724	174824

ical and observed frequencies is (4 - 1) (4 - 1) + 1 or 10 The value of X^2, calculated from the data in Table IV, is 1 3171. The corresponding value of P, which is the probability that as bad, or worse, an agreement between observed and theoretical frequencies could occur from the fluctuations of random sampling is, according to Pearson's Tables,[1] 0 996911 or almost certainty The percentage errors were, therefore, distributed in such a manner as to be uncorrelated with the values of X at which they were used

The equation, $Y = AX^2$, was fitted to the hypothetical set of data in Table III by the method of least squares as ordinarily described. If AX^2 represents a calculated value of the dependent variable and Y represents the corresponding observed value, the difference between these two values is $AX^2 - Y$ and the square of the difference is $A^2X^4 - 2AX^2Y + Y^2$ The sum of the squares of all the differences is $A^2\Sigma X^4 - 2A\Sigma X^2Y + \Sigma Y^2$ The value of this expression will be a minimum when its derivative with respect to A is equal to zero. Differentiating and equating to zero yields the following equations for the determination of A:

$$2A\Sigma X^4 - 2\Sigma X^2 Y = 0 \tag{1}$$

$$A = \frac{\Sigma X^2 Y}{\Sigma X^4} \tag{2}$$

The value of A calculated from the data in Table III by means of equation (2) is 100 6250.

If residuals of the type suggested by Pearl and Reed are employed, A is calculated as follows Let AX^2 represent a calculated value of the dependent variable, as before, and let Y represent the corresponding observed value Then the difference between the two values, expressed as a fraction of the observed

[1] Pearson, Karl Tables for Statisticians and Biometricians. Cambridge University Press, London (1924).

TABLE IV

Chi-square test for contingency applied to the distribution of the percentage errors of measurement. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Error (Per cent)				
	0.0 to ±1.9	±2.0 to ±3.9	±4.0 to ±5.9	±6.0 and over	Total
1 to 10	2 (2.1951)	4 (3.9024)	2 (2.4390)	2 (1.4634)	10
11 to 20	2 (2.1951)	4 (3.9024)	3 (2.4390)	1 (1.4634)	10
21 to 30	2 (2.1951)	4 (3.9024)	2 (2.4390)	2 (1.4634)	10
31 to 41	3 (2.4146)	4 (4.2926)	3 (2.6829)	1 (1.6097)	11
Total	9	16	10	6	41

$X^2 = 1.3171$

$n' = 10$

$P = 0.996911$

value, is $\frac{AX^2-Y}{Y}$ or $\frac{AX^2}{Y}-1$. The square of this relative deviation is $\frac{A^2X^4}{Y^2}-\frac{2AX^2}{Y}+1$ and the sum of the squares of the 41 relative deviations is $A^2\Sigma\frac{X^4}{Y^2}-2A\Sigma\frac{X^2}{Y}+41$. This expression will likewise have its minimum value when its derivative with respect to A is equal to zero. Differentiating and equating to zero, as before, leads to the following equations for the determination of A:

(3) $$2A\Sigma\frac{X^4}{Y^2}-2\Sigma\frac{X^2}{Y}=0$$

(4) $$A=\frac{\Sigma\frac{X^2}{Y}}{\Sigma\frac{X^4}{Y^2}}$$

Applying equation (4) to the given set of data gives a value of 99.7573 for A. This value of A is closer to the true value 100, than the value which was calculated by means of equation (2) but the improvement was not as great as might be expected

It occurred to the writer that if the deviations of the calculated, from the observed, values of the dependent variable were expressed as fractions of the calculated values, a more accurate value of A could be obtained.

The relative deviation expressed in this manner is $\frac{AX^2-Y}{AX^2}$ or $1-\frac{A^{-1}Y}{X^2}$. The square of this deviation is $1-\frac{2A^{-1}Y}{X^2}+\frac{A^{-2}Y^2}{X^4}$ and the sum of the squares of the 41 relative deviations is $41-2A^{-1}\Sigma\frac{Y}{X^2}+A^{-2}\Sigma\frac{Y^2}{X^4}$. Differentiating this expression with respect to A and equating to zero yields the following equations for the determination of A:

(5) $$2A^{-2}\Sigma\frac{Y}{X^2}-2A^{-3}\Sigma\frac{Y^2}{X^4}=0$$

(6) $$A=\frac{\Sigma\frac{Y^2}{X^4}}{\Sigma\frac{Y}{X^2}}$$

The value of A, calculated from the data by means of equation (6), is 100.1210 which is nearer to the true value than either of the values calculated by the two preceding methods. However, it is evident that equation (6) failed to give results as precise as one would expect, in view of the method by which the observed values of Y were obtained.

The reason for this discrepancy can be made most apparent by returning to the analogy existing between the application of the method of least squares to curve fitting and the calculation of the arithmetic mean of a number of measurements of a single, constant quantity

Let $m_1, m_2, m_3, \ldots m_n$ represent measured values of the same constant quantity and let their arithmetic mean be represented by M. If each measurement is divided by the arithmetic mean of all the measurements, the resulting distribution of these relative values will be normal if the original measurements were distributed normally. The arithmetic mean of these relative values will obviously be unity.

Let $\frac{m_1}{M}, \frac{m_2}{M}, \frac{m_3}{M}, \ldots \frac{m_n}{M}$ represent the relative values of the measurements. The arithmetic mean of these values is unity. Therefore, the deviation of any relative value, $\frac{m}{M}$, from the mean is $1 - \frac{m}{M}$.

Let it be assumed that the value of the arithmetic mean of the original measurement, M, is unknown and is represented by Z. Then any measurement, m, expressed as a fraction of Z, is $\frac{m}{Z}$. According to the discussion in the two preceding paragraphs, it might appear that Z must have such a value that the sum of the squares of the deviations, $1 - \frac{m}{Z}$, is a minimum. However, this is not the case. It may be demonstrated that the value of the expression $\Sigma\left(1 - \frac{2m}{Z} + \frac{m^2}{Z^2}\right)$, is a minimum when Z has some other value than the arithmetic mean of the original measurements. The sum of the squares of the residuals may be written, $n - 2Z^{-1}\Sigma m + Z^{-2}\Sigma m^2$. Differentiating this expression with respect to Z and equating to zero yields the following

equations for the determination of Z :

(7) $$2Z^{-2}\Sigma m - 2Z^{-3}\Sigma m^2 = 0$$

(8) $$Z = \frac{\Sigma m^2}{\Sigma m}$$

The value of Z, calculated by means of equation (8), is obviously not the arithmetic mean of the original measurements. The fallacy in the deduction of this equation is readily apparent

Instead of using residuals of the type, $1 - \frac{m}{Z}$, and differentiating the sum of the squares of the residuals with respect to Z, one should use residuals of the type, $V - \frac{m}{Z}$, in which V represents the arithmetic mean of the relative values, $\frac{m}{Z}$, of the measurements The sum of the squares of the residuals should be differentiated with respect to V. The square of the residual, $V - \frac{m}{Z}$, is $V^2 - \frac{2m}{Z} + \frac{m^2}{Z^2}$ and the sum of the squares of all the residuals may be written $nV^2 - \frac{2V}{Z}\Sigma m + \frac{1}{Z^2}\Sigma m^2$ Differentiating with respect to V and equating to zero yields the following equations for the determination of V.

(9) $$2nV - \frac{2}{Z}\Sigma m = 0$$

(10) $$V = \frac{\frac{1}{Z}\Sigma m}{n}$$

Since the value of V is known to be unity, equation (10) may be written:

(11) $$n = \frac{1}{Z}\Sigma m$$

from which Z may be readily calculated as follows

$$(12) \qquad Z = \frac{\Sigma m}{n}$$

Equation (12) is obviously nothing more than the simple formula for the calculation of the arithmetic mean of the original measurements, which is sufficient evidence that the reasoning involved in its deduction is sound.

It is now readily apparent why equation (6) did not yield results which were consistent with the data in Table III. The ratio, $\frac{Y}{AX^2}$, is analogous to the ratio, $\frac{m}{Z}$, and residuals of the type, $V - \frac{Y}{AX^2}$, should have been used in fitting the equation instead of residuals of the type, $1 - \frac{Y}{AX^2}$.[1] The square of the residual, $V - \frac{Y}{AX^2}$, is $V^2 - \frac{2VY}{AX^2} + \frac{Y^2}{A^2X^4}$. The sum of the squares of the 41 residuals is $41V^2 - \frac{2V}{A}\Sigma\frac{Y}{X^2} + \frac{1}{A^2}\Sigma\frac{Y^2}{X^4}$. Differentiating this expression with respect to V and equating to zero yields the following equations for the determination of V:

$$(13) \qquad 82V - \frac{2}{A}\Sigma\frac{Y}{X^2} = 0$$

$$(14) \qquad V = \frac{\frac{1}{A}\Sigma\frac{Y}{X^2}}{41}$$

Substituting the known value, unity, for V in equation (14) yields the following equations for the determination of A.

$$(15) \qquad 41 = \frac{1}{A}\Sigma\frac{Y}{X^2}$$

$$(16) \qquad A = \frac{\Sigma\frac{Y}{X^2}}{41}$$

[1]Residuals of the type, $\frac{AX^2}{Y} - 1$, are analogous to those of the type, $\frac{Z}{m} - 1$, which also lead to incorrect results.

Applying equation (16) to the data in Table III gives A a value of 100.0000, which coincides exactly with the true value from which the data were originally calculated. Equation (16) was, therefore, the correct equation to use in interpreting the data given in Table III Although the use of residuals of the types, $\frac{AX^2}{Y} - 1$ and $1 - \frac{Y}{AX^2}$, gave better approximations to the true values of A than the use of the simple residuals, $AX^2 - Y$, neither of the two gave results which were entirely in accord with the derivation of the data

Yule[1] suggested that the geometric mean might often prove useful in comparing the frequency distributions of different sets of data, in which the dispersion of the individual measures about their means was influenced by the magnitude of the means It appeared to the writer that the use of residuals of the type, $\log AX^2 - \log Y$, might give a good approximation to the true value of A in fitting the given equation. It is evident that the ratio, $\frac{AX^2}{Y}$, approaches unity as the residual, $\log AX^2 - \log Y$, approaches zero.

This logarithmic residual may be written, $\log A + 2 \log x - \log Y$, and its square is $(\log A)^2 + 4(\log X)^2 + (\log Y)^2 + 4(\log A)(\log X) - 2(\log A)(\log Y) - 4(\log X)(\log Y)$. The sum of the squares of the 41 residuals is $41(\log A)^2 + 4\Sigma(\log X)^2 + \Sigma(\log Y)^2 + 4(\log A)\Sigma(\log X) - 2(\log A)\Sigma(\log Y) - 4\Sigma(\overline{\log X} \cdot \overline{\log Y})$. Differentiating this expression with respect to $\log A$ and equating to zero yields the following equations for the determination of A :

$$(17) \qquad 82(\log A) + 4\Sigma(\log X) - 2\Sigma(\log Y) = 0$$

$$(18) \qquad \log A = \frac{(\log Y) - 2\Sigma(\log X)}{41}$$

---

[1]Loc. cit.

The value of $\log A$, calculated from the given set of data by means of equation (18), is 1.9997369, which gives A a value of 99.9394. This value of A comes closer to the true value than those calculated by means of residuals of the types, $1 - \frac{Y}{AX^2}$ and $\frac{AX^2}{Y} - 1$. However, since the use of the geometric mean is not rigorously justified when the distribution of the measures about the arithmetic mean is symmetrical, the use of logarithmic residuals in curve fitting can not give precise results when the errors of measurement are distributed as they were in the given set of data.

In any application of the method of least squares to a practical problem, the procedure of the investigators should be governed by the nature of the data to which it is being applied. In many instances the correct procedure can be deduced by a careful consideration and evaluation of the accuracy of the methods of measurement used in obtaining the data. Unfortunately, however, some sources of error are not always readily apparent at the time the data are collected, and occasionally can not be quantitatively estimated even though they are known to exist. If the nature of the mathematical relationship existing between the dependent and independent variables is known, all that remains is to find the most probable values of the constants in the equation

A statistical study of the deviations of the observed values of the dependent variable from the corresponding calculated values, obtained after fitting the equation by several different methods, may be of much help in deciding which method of fitting was most consistent with the nature of the data. For example, Table V gives the results of applying the chi-square test for contingency to the distribution of the deviations of the observed values of Y from the calculated values obtained when residuals of the type, $AX^2 - Y$, were used in fitting the equation, $Y = AX^2$, to the data in Table III. The value of P is only 0.005061 and a mere inspection of the table itself shows that large deviations tend to occur more frequently, and small deviations less frequently, as

TABLE V

Chi-square test for contingency applied to the distribution of the deviations of the type, AX^2-Y. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Deviation				
	10 to ± 1999	± 2000 to ± 3999	± 4000 to ± 5999	± 6000 and over	Total
1 to 10	10 (7.3171)	0 (1.2197)	0 (0.9756)	0 (0.4878)	10
11 to 20	10 (7.3171)	0 (1.2197)	0 (0.9756)	0 (0.4878)	10
21 to 30	6 (7.3171)	1 (1.2197)	3 (0.9756)	0 (0.4878)	10
31 to 41	4 (8.0488)	4 (1.3415)	1 (1.0732)	2 (0.5366)	11
Total	30	5	4	2	41

$\chi^2 = 23.5989$

$n' = 10$

$P = 0.005061$

the values of X increase. If the true nature of the values of Y in Table III were not known in advance, this distribution of the deviations would be sufficient evidence that the method of fitting the equation was not consistent with the accuracy of the measurements made when the data were collected.

Tables VI, VII, and VIII give, respectively, the distributions of the deviations of the types, $\frac{AX^2}{Y} - 1$, $1 - \frac{Y}{AX^2}$, and $\log AX^2 - \log Y$, when the corresponding residuals were used in fitting the equation.[1] The value of P is high in each case, indicating that, although the use of residuals of these types did not give results which were precisely accurate, nevertheless, they yielded values of A which were well within the limits of the probable error to be expected in any practical investigation.

As a matter of fact, this is a rather fortunate circumstance, since the only method of fitting the equation given above which yielded exactly the correct value of A cannot be applied to fitting an equation containing more than one undetermined constant. The applicability of residuals of the types, $1 - \frac{Y}{f(X)}$ and $\log f(X) - \log Y$ is also somewhat limited. However, any equation which can be fitted by the method of least squares at all can still be fitted when residuals of the type, $\frac{f(X)}{Y} - 1$, are employed.

SUMMARY AND CONCLUSIONS

The method of least squares can be a more valuable tool in statistical work when the fundamental theory upon which the method is based is taken into consideration. The use of residuals of the type, $f(X) - Y$, is probably justified in fewer practical

---

[1]The distribution of the deviations obtained when the equation was fitted to the data by means of equation (16) is identical with the distribution of the errors given in Table IV.

problems than the use of residuals of some other form. The type of residual to be employed should be governed by the nature of the data to which the method of least squares is being applied.

The use of relative residuals of the type suggested by Pearl and Reed may be of much value in many instances but will not give results which are precisely accurate, even though the distribution of the percentage errors of measurement is strictly normal. The results can be improved by expressing the deviations of the observed from the calculated values of the dependent variable as fractions of the calculated, rather than the observed, value.[1]

The use of logarithmic residuals may give more accurate results than the use of residuals of the type suggested by Pearl and Reed, even though the distribution of the percentage errors of measurement is normal.

The chi-square test for contingency may be of much help in selecting the type of residual most consistent with the errors of measurement made in obtaining the data when sufficient information regarding the accuracy of the measurements is not available.

---

[1]Residuals of this type have been used by Hendricks, Lee, and Titus at the U. S. Animal Husbandry Experiment Farm, Beltsville, Maryland, in the fitting of growth curves.

Hendricks, W. A., A. R. Lee, and H. W. Titus. Early growth of White Leghorns, Poultry Sci. 8 (6); pp. 315-327 (1929).

Titus, H. W., and W. A. Hendricks. The Early Growth of Chickens as a Function of Feed Consumption Rather Than of Time. Conference Papers of the Fourth World's Poultry Congress, Section B (Nutrition and Rearing): pp. 285-293 (1930).

The use of such residuals leads to results which appear to give a better description of the data than when simple residuals of the type, $f(X) - Y$, are employed.

TABLE VI

Chi-square test for contingency applied to the distribution of the deviations of the type, $\frac{AX^2}{Y} - 1$. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Deviation				
	0 000 to ± 0 019	± 0.020 to ± 0.039	± 0 040 to ± 0 059	± 0 060 and over	Total
1 to 10	4 (4.1463)	3 (3.1707)	2 (1.7073)	1 (0.9756)	10
11 to 20	4 (4.1463)	4 (3.1707)	1 (1.7073)	1 (0.9756)	10
21 to 31	4 (4 1463)	3 (3.1707)	1 (1.7073)	2 (0.9756)	10
31 to 41	5 (4.5610)	3 (3.4878)	3 (1.8780)	0 (1.0732)	11
Total	17	13	7	4	41

$\chi^2 = 3.8182$

$n' = 10$

$P = 0.921027$

TABLE VII

Chi-square test for contingency applied to the distribution of the deviations of the type, $1 - \frac{Y}{AX^2}$. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Deviation				
	0 000 to ±0.019	±0.020 to ±0.039	±0.040 to ±0.059	±0.060 and over	Total
1 to 10	3 (3.9024)	4 (3.4146)	2 (1 7073)	1 (0.9756)	10
11 to 20	4 (3.9024)	3 (3.4146)	2 (1.7073)	1 (0.9756)	10
21 to 30	4 (3 9024)	3 (3.4146)	1 (1 7073)	2 (0.9756)	10
31 to 41	5 (4.2927)	4 (3.7561)	2 (1 8780)	0 (1.0732)	11
Total	16	14	7	4	41

$\chi^2 = 3.0984$

$n' = 10$

$P = 0.959091$

TABLE VIII

Chi-square test for contingency applied to the distribution of the deviations of the type, log AX^2 - log Y. The theoretical frequencies for each compartment are given in parentheses.

Value of X	Magnitude of Deviation				
	0.000 to ±0.009	± 0.010 to ±0.019	± 0.020 to ±0.029	± 0.030 and over	Total
1 to 10	6 (6.0976)	2 (2.4390)	2 (0.9756)	0 (0.4878)	10
11 to 20	6 (6.0976)	3 (2.4390)	0 (0.9756)	1 (0.4878)	10
21 to 30	6 (6.0976)	2 (2.4390)	1 (0.9756)	1 (0.4878)	10
31 to 41	7 (6.7073)	3 (2.6829)	1 (1.0732)	0 (0.5366)	11
Total	25	10	4	2	41

$X^2 = 4.4989$

$n' = 10$

$P = 0.872945$

EDITOR'S NOTE

It is with great pleasure that the *Annals* brings to its readers information concerning the *Nordic Statistical Journal*, edited by Dr. Thor Andersson. This publication is of great merit, and the work of its contributors compares very favorably with that found in *Biometrika* and *Metron*. Americans will do well to study carefully the contributions which Scandinavians are making to statistical methodology.

Nordic Statistical Journal

EDITED BY

THOR ANDERSSON

VOLUME 1

480

ARTICLES IN NORDISK STATISTISK TIDSKRIFT.

Vol 1

STATISTIKISERING, Brev till John Burns från .... UTGIVAREN
STATISTICIZATION, Letter to John Burns from ... . THE EDITOR
DIE VARIATIONSBREITE BEIM GAUSSSCHEN FEHLERGESETZ, I — L. v. BORTKIEWICZ
DAS GESETZ DER GROSSEN ZAHLEN UND DER STOCHASTISCH-STATISTISCHE STANDPUNKT IN DER MODERNEN WISSENSCHAFT . ... . AL. A. TSCHUPROW
STATISTICS AND PREHISTORIC SCIENCE .... O. MONTELIUS
BIOLOGI OG STATISTIK .... W JOHANNSEN
STATISTIK OG HISTORIE .... A. O JOHNSEN
DEN ISLANDSKE STATISTIKS OMFANG OG VILKAAR., THORSTEIN THORSTEINSSON
BEFOLKNINGSSTATISTIKEN I FINLAND REORGANISATIONSPLANER — A. E. TUDEER
NORDMANNEN I VÄRLDEN .... THOR ANDERSSON
JORDBRUKETS UTVECKLING I VISSA DELAR AV SKÅNE OCH DANMARK — ERNST HÖIJER
DIE ALLRUSSISCHEN LANDWIRTSCHAFTSZAHLUNGEN VON 1916 UND 1917 — STAN KOHN
INTERSKANDINAVISK HANDELSSTATISTIK 1912—1918 .... JOHS. DALHOFF
LEHRBÜCHER DER STATISTIK .... AL. A. TSCHUPROW
DIE VARIATIONSBREITE BEIM GAUSSSCHEN FEHLERGESETZ, II — L. v. BORTKIEWICZ
STATISTISKA SAMFUNDET I FINLAND . .. . A. E. TUDEER
DEN NORSKE ÖVERSJÖISKE UTVANDRING .... E. STORSTEEN
SVENSKA JORDENS ÄGARE OCH BRUKARE .... PAUL DAHN
LEHRBÜCHER DER STATISTIK ... . , AL. A. TSCHUPROW
IST DIE NORMALE STABILITÄT EMPIRISCH NACHWEISBAR AL. A. TSCHUPROW
ON THE EFFECTIVITY OF WEATHER WARNINGS .... A. ANGSTRÖM
RIKSSTATISTIKENS CENTRALISERING I AMERIKAS FÖRENTA STATER
ET FOLKEREGISTER I DANMARK . . . . JOHS DALHOFF
DER EINFLUSS DES KRIEGES AUF DIE GEBURTEN .... E CZUBER

Vol 2

WAHRSCHEINLICHKEIT UND STATISTISCHE FORSCHUNG NACH KEYNES — L v. BORTKIEWICZ
AUFGABEN UND VORAUSSETZUNGEN DER KORRELATIONSMESSUNG — AL A TSCHUPROW
SOCIALSTATISTIKENS CENTRALISERING OCH SOCIALSTYRELSENS INDRAGNING I FINLAND ... . THOR ANDERSSON
FORHOLDET MELLEM KJØNNENE I DEN STAAENDE BEFOLKNING OG SEKSUALPROPORSJONEN FOR DE FØTTE .. .. INGVAR WEDERVANG
DET SVENSKA FÖDELSEÖVERSKOTTETS UTKOMSTMÖJLIGHETER I EGET LAND — FR SANDBERG
EIN BÜRGERLICHER HAUSHALTUNGSAUFWAND . . .. E. CZUBER
SVERGES HANDELSSTATISTIK OCH DE STATISTIKSAKKUNNIGA THOR ANDERSSON
SAMFÄRDSELNS PERIODICITET .. .. Y. NYLANDER
BUSINESS STATISTICS .... AL. A TSCHUPROW
FOLKREGISTRERINGEN I NORGE . ... G. AMNÉUS
FOLKOMRÖSTNINGEN DEN 27 AUGUSTI 1922 ANGÅENDE RUSDRYCKSFÖRBUD — OTTO GRÖNLUND
RIKSSTATISTIKENS CENTRALISERING I FINLAND .... A. E. TUDEER
RIKSSTATISTIKENS CENTRALISERING I CANADA . ... THOR ANDERSSON
ZWECK UND STRUKTUR EINER PREISINDEXZAHL, I ... . L. v. BORTKIEWICZ
ARBEIDSBESPARENDE METODER I STATISTIKKEN ... . ADOLPH JENSEN
OM MIDDELFEJLEN VED PARTIELLE UNDERSØGELSER .. . HANS CL. NYBØLLE
FORSLAG TIL CIVILSTANDSREGISTRERING I NORGE .... G. AMNÉUS
KØBENHAVNS FOLKEREGISTER ... . BERTEL DAHLGAARD
VÄXTODLINGEN I SVERGE .... , ... , ..., ... ERNST HÖIJER
INTERSKANDINAVISK HANDELSSTATISTIK 1912—1922 .... JOHS DALHOFF

Vol 3

DET INTERNATIONALE STATISTISKE INSTITUTS MØDE I BRUXELLES I OKTOBER 1923 .... ADOLPH JENSEN
DANSKE STATISTIKERES F[illegible] . H. HØST
STANDSREGISTRERINGEN I [illegible] AMNÉUS
JERNKONTORET OCH BERG[illegible]
STUDIER I SVENSK ALKOHOLSTATISTIK 1, 2 .... HANS GAHN
GRUNDBEGRIFFE UND GRUNDPROBLEME DER KORRELATIONSTHEORIE — AL. A. TSCHUPROW
ZWECK UND STRUKTUR EINER PREISINDEXZAHL, II... .. L. v. BORTKIEWICZ
THE FOREST RESOURCES OF SWEDEN .... TOR JONSON
THE ORE RESOURCES AT THE KIIRUNAVAARA AND GELLIVARE MINES — WALFR. PETERSSON
SVERGES POSTVÄSEN 1620—1924 . ... Y. NYLANDER
STATISTICS OF INDUSTRIAL PRODUCTION ... ... ADOLPH JENSEN
EN BOK OM KOOPERATIONEN .... CURT ROTHLIEB
ZIELE UND WEGE DER STOCHASTISCHEN GRUNDLEGUNG DER STATISTISCHEN THEORIE ... . AL A TSCHUPROW
ZWECK UND STRUKTUR EINER PREISINDEXZAHL, III ...... L v. BORTKIEWICZ
FEIL I DET BEFOLKNINGSSTATISTISKE MATERIALE . . . . . HENRIK PALMSTRÖM
UNDERSÖKNING RÖRANDE DEN ANIMALISKA PRODUKTIONENS STORLEK I SVERGE, I . . . . ERNST HÖIJER
THE PROSPECTS OF THE PAPER INDUSTRY ... . HANS ANSTEIN
STATS- OG KOMMUNEREGNSKABERNE I DE NORDISKE LANDE CHRISTIAN OLSEN

Vol 4

SVENSKA FÖRSÄKRINGSFÖRENINGEN OCH STATISTIKEN ... . THOR ANDERSSON
WILHELM LEXIS UND SEINE BEDEUTUNG FÜR DIE VERSICHERUNGSWISSENSCHAFT . ... W. LOREY

Nordisk Statistisk Tidskrift started in 1922. It is chiefly written in Nordic tongues. There are also published articles in English and German. To some articles in Nordic there are summaries in English or German. Now the chance is taken to realize the original scheme of publishing two editions, one in Nordic tongues and the other in English. The edition in non-Nordic tongues is published in English also because of the fact that the millions of descendants of the Nordic peoples, now living beyond the boundaries of the Nordic states, are mainly working in English-speaking countries.

Nordic Statistical Journal has five departments. articles, reviews of books, minor communications, bibliographical lists of Nordic statistics, and recent periodicals and new books. In general, all departments will be represented in every number.

Nordic Statistical Journal is published quarterley, the four numbers making a volume of about 640 pages. The subscription rate for a volume — post free — is 30 Swedish crowns

Subscriptions may be sent to Nordic Statistical Journal, Stockholm, Sweden.

The subscription rate through booksellers is 35 Swedish crowns.

Editorial communications and all publications should be adressed to THOR ANDERSSON, Dr. Ph., Stockholm, Sweden.

Aftonbl. tr., Sthlm 1929.

Nordic Statistical Journal

EDITED BY

THOR ANDERSSON

VOLUME 2 PARTS 1 & 2

REPRINT AND TRANSLATION FROM NORDISK FÖRSÄKRINGSTIDSKRIFT 1930.

Nordic Statistical Journal. Volume 1. Edited by THOR ANDERSSON. Stockholm 1929. Pp. 639. Reviewed by Dr. phil. CARL BURRAU

We, the inhabitants of the Nordic countries, are perhaps somewhat inclined to take a certain inner pride in our — as it seems to us — high civilization and to attach still more importance to ourselves in this respect during the later years, when the "Ragnarok" of the great war had devastated most of the other civilized countries and handicapped them in their competition with us. Let us hope that there are some good grounds for our selfsatisfied opinion! It is not difficult to find some facts indicating that we are right in this self-respect, even if we go to the very summits of civilization — let us think of the "Acta matematica", for instance. But if we are right, it may be very necessary for us to be on our guard against the danger of stagnation, of the standstill, where we begin to lull ourselves into the pleasant dream that our position is unshakeable, and that we may now repose on our laurels. Therefore, we must honour the persons who do not allow us to go to rest, the persons who spur us on to do our very best

Thor Andersson is one of those whom we must honour for such an influence In the field of statistics he seeks to be our scientific conscience. He swings his whip over our heads mercilessly and drives to activity everybody who is able to produce something, however small or great, within the field of statistics. But he is not content with that! He is not content with the achievement of having filled a long and imposing row of volumes of the "Nordisk Statistisk Tidskrift" with valuable essays and treatises written by Scandinavian as well as by leading foreign authors — all the non-scandinavian countries are now to see and feel the warmth of the light from the North. His journal is now to become an inter-

national publication, but still with an indication of its Nordic origin in its title. The first volume of the "Nordic Statistical Journal" — simultaneously forming the 8th volume of the original journal — has appeared. And it is not a trifling thing, this volume of 639 pages in great octavo! It is great in its composition, still more soaring in its purposes and ends for the future, and promising, when we consider what "the man at the wheel" has collected in these 639 pages by means of an unusual perseverance in unfailing love for the task and in spite of many — too many — external adversities

The leading thought of the work is the same as, now soon a decennium ago, led Thor Andersson to found the *Nordisk Statistisk Tidskrift*. It is a child of the Greeks' idea of chaos and cosmos, or rather a consequent, modern continuation of this idea. Statistics is the most important means for bringing our existence over from chaos to cosmos. Statistics acquaints us with the real circumstances, and the knowledge, the real knowledge of the things, will then show how to bring things in their right places, so that the entirety becomes the arranged cosmos. But there is still much to do! We have not yet been able to elevate statistics to the rank of an observing natural science it should have, to be able to give us the real science of the things, alluded to above. In 1922 the thought was to be in the front-rank in the work for this purpose. And we have to be obliged to Thor Andersson for the strenuous work he has performed for his idea during the past years, and now it will be done on a still broader basis, i. e. for an international public, yet under Nordic leadership.

Let us study a little more closely how this new volume I seeks to perform its work in the service of the mentioned idea.

With, in a good meaning, a journalistic feeling for actualities the volume appears as a sort of jubilee-gift to *Bortkiewicz* on his sixtieth anniversary and it is therefore opened by a good full-page picture of this scientist who has given so valuable contributions to the original journal. To the reader, the following essays seem to arrange themselves into three groups which — just in order to give a name to the special groups — could be designated as *olden times*, *present times*, and *future times*.

In the group belonging to the olden times, the editor seeks to show how deeply rooted the statistical science is in the Nordic peoples by introducing a number of great men of Nordic origin, each in his way, a pioneer. These men are represented partly in full page pictures, partly in the text It may not surprise us that none of them is a "professional statistician", for the profession is only now being created. But they belong, each in his way, to the founders of this branch. In the eightteenth century *Wargentin*, the astronomer, founded population statistics which is of fundamental importance to demographics. The essay on him is particularly well written by *Nordenmark*

The memory of the now nearly forgotten *Eilert Sundt*, who, by his activity as a clergyman, was brought to make scientifical investigations of the society where he lives, and who gradually becomes a social-statistician of high rank, is revived both by a reprint of his peculiar essay of 1858. "On Piperviken and Ruselokbakken (Investigations of the conditions and morals of the working-class in Christiania)" and by a scientific estimation of him ("Eilert Sundt's law") by *Rygg* who also gives an instructive account of how Sundt was disfavoured by his contemporaries, naturally in the first place by the politicians who had to do with the granting of money for his investigations! Unfortunately the politicians of the present times are not better; about that Thor Andersson himself could write a sad chapter!

Then follows *Thiele*, whose principal scientific passion, "the theory of observations", is simply the foundation of what is now more generally called mathematical statistics, and finally *Johannsen*, the investigator of heredity, who is commemorated by a picture as well as by a reprint and a translation into English of his contribution to the first volume of the original journal: "Biology and statistics".

Several other essays like those mentioned, also belong to the olden times. Thus *H. Palmströms's* essay on the first census in Norway in 1769 and that of *Thorsteinsson* on the census of Iceland in 1703

The essays which the reader naturally refers to the "present times" are evidently caused by the editor in order to show the non-Scandinavian world the conditions in the

Nordic countries in two respects, both extremely important from a statistical point of view: the population registration and the industries, thus, firstly, how we gain our knowledge about the number and the composition of the population, and, secondly, how these people support themselves.

The editor could not have found any person more fit to write the "general" article about population registration than *Amnéus*, the director of the Oslo population register, whose institution is up to the standard and also has served as a model in many places, among others in Denmark. The condition of these matters in the different countries is further treated by the editor as far as Sweden is concerned, and as to Denmark by not less than two authors, *Bonde* and *Dalgaard*, and with regard to Finland by *Kovero*. These are very instructive essays which illustrate the importance of these things in the right way. One learns how even the "torso" (a not unjustified epithet for the arrangement introduced in Denmark, which was originally excellently planned, but which has been more than half-way broken to pieces by smallminded and short-sighted politicians, of course under the pretext of economy) of a population register, as that of Denmark, thanks to the fact that it is obligatory, gives an excellent support in many ways, among others for the 5-year censuses. One learns how deplorably far behind matters are in Wargentin's native country, where it was naturally necessary to perform registration in the large towns, but how the acclomplishment of the work is hazarded by the rather antiquated and burdensome obligatory collaboration with the clergy — and by still many other things. A survey like this, presented to an international audience is perhaps more than anything else suited to advance the population register movement, which *shall*, however, once triumph by its inner necessity.

Next the editor has intended to give a picture of the industrial statistics of the Nordic states. He has himself undertaken to treat the most important part: "the mother industry", agriculture.

Aage I. C. Jensen treats fishery, *Sköien* shipping, *Geijer* the ore resources of the Nordic countries, *Velander* the water powers of the Nordic countries, and, forestry, finally, is treated by *Aminoff* (Sweden), *Sandmo* (Norway), and *Cajander* (Finland).

To the "present times" we may also count *Storsteen's*, to us, the inhabitants of the Nordic capitals very interesting article on "The expenses of living in the Nordic capitals", a subject full of pitfalls for a less experienced statistician, but here treated with excellent fineness and with a clear prescience of the difficulties.

To the same group belongs *Thorberg*· Statistics and trade-union movement", a rather short but extremely interesting essay, not least on account of the author's position as president of the national organisation of the Swedish trade-unions. Here fall the weighty words about the social-political institution erected by the League of Nations for international labour organisation, that "the work of this organization is rendered extremely difficult by the fact that it has hitherto been almost impossible to arrive at any comparability between the statistics of the different countries"

Finally there is *Lander's*. "Some remarks on the income statistics of the census in 1920", which, according to its title, seems to belong to the present times, but which, according to its contents, is in the first place, a scientifical arithmetical example for the illustration of the applicability of Pareto's law, concluding in some wishes with regard to the future official investigations of income. This essay can therefore be said to form the transition to the last group.

When I have permitted myself to designate the third group of essays as "future", there may thereby, as a matter of fact, not be understood any paradoxical possibility of prophesying the statistics of the future. I have only wished to emphasize the editor's desire that his journal may also be one of the laboratories where the instruments for the treatment of the future statistics are created. This side of the matter has always had the editor's supreme interest; you may think only of the contributions to the previous volumes, which Bortkiewicz, Tschuprow and others have brought. We can call this side the theoretical or perhaps the mathematic-statistical one It has been an urgent need in this volume I to show, that also we, in the Nordic countries think of this side of the matter, and among the authors of the six essays of which this group consists (besides the above mentioned contribution

by Linders) we also find all the four Nordic countries represented.

Although the chief importance of treatises of this kind would seem to fall within the realms of theory, still one of this essays, namely *Nybölle's* "Interpolation in statistics" is of rather eminent practical importance and use. Beside — or rather because of — his clear and sharp differentiation between the purely mathematical and the statistical meaning of the word interpolation and the very near connection of the last mentioned notion with that of adjustment, he here gives exclusively practical advice and instructions useful in circumstances which, so to say, belong to the every day life of the statistician. This treatise will be very welcome to many colleagues. In some opposition hereto stands *Ragnar Frisch*, "Correlation and scatter in statistical variables", in size as well as in importance one of the biggest treatises of the volume which will prick the conscience of many as it will make them clearly feel the obligation to penetrate more deeply into its contents — the author himself namely tells us that he has come "to various results, some of which are known, and others which are new, so far as I am aware" — but will easily feel frightened by the author's imposing mathematical apparatus, which is nothing less than n dimensional vectors and appertaining matrices and orthogonal transformations. But one ought not to be frightened by these heavy implements. And this so much the less as the author presupposes no elementary knowledge in the field of vector calculation On the contrary, he explains his whole apparatus thoroughly. There is no need of having heard words as vector or orthogonal before, and one will still be able to study this work, which, on this account has naturally become somewhat extensive (67 pages). It may be greeted with great satisfaction that one thus begins to attack the problems of theoretical statistics with such weapons. Terms such as scatter and correlation are so fundamental to science that we cannot take into use an apparatus precious enough to attack the problems contained in these terms. Here the best is not too good.

Further we meet *Jerneman*: "To the method of sampling" and *Lindeberg:* "Some remarks on the mean error of the per-

centage of correlation", essays well suited to waken respect for Nordic science abroad.

The contact with the kindred science, social economics, is in the volume attended to by *Gjermoe* "The amplitude of industrial fluctuations", an essay covering 63 pages, which struggles with the difficult and not yet very well defined notions: times of ascent and descent, crises, fluctuations, and so on. The notion of "trend", so prominent in all the ultra modern investigations, belongs to the here used apparatus, and in a quotation — this essay is very abundant in quotations, which will be praiseworthy — we learn that it was already found by Hooker in 1901 and was defined by him as "the direction in which the variable is really moving when the oscillations are disregarded"

As rather specially belonging to the "future" we meet finally *Ångström's* brilliant essay on meteorology and statistics and have a presentiment that the latter is destined to play once the principal rôle before the former.

We will not end this review without another congratulation to Thor Andersson for having brought forth this volume I, followed by the wish and hope that his ideal struggle for the highest aims will meet with the wished for success before he grows too tired to fight against adversity It is sadly known that the Swedish Riksdag has not granted him the necessary subsidy in spite of the fact that such recommendations as the following one could be appended to the petition

"By the way in which Dr. Thor Andersson has edited Nordisk Statistisk Tidskrift, he has, according to our opinion, rendered great services towards the advancement of scientifical statistics and towards the spread of the knowledge of its extraordinary importance, not only for other sciences, but also, and not least, for the obtaining of a real knowledge of the social and economical structure of society, a knowledge that is necessary if the public measures of correcting social evils and of furthering industry will have the wished for effect. His name guarantees that the journal will, also in the future, hold the same prominent position as it now holds among publications of this kind

Stockholm January 17th, 1929.

E Phragmén *P G. Laurin*"